HYBRID STOCHASTIC MODELS FOR REMAINING LIFETIME PROGNOSIS

DISSERTATION

Steven M. Cox, Major, USAF

AFIT/DS/ENS/04-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

Research sponsored primarily by the Air Force Office of Scientific Research and secondarily by the Air Force Research Laboratory, Human Effectiveness Directorate, Air Force Materiel Command, USAF. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation thereon. The views and conclusions contained in this dissertation are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the Air Force Research Laboratory, Department of Defense, or the United States Government.

AFIT/DS/ENS/04-01

HYBRID STOCHASTIC MODELS FOR REMAINING LIFETIME PROGNOSIS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Steven M. Cox, BS, MS

Major, USAF

August 2004

AFIT/DS/ENS/04-01

HYBRID STOCHASTIC MODELS FOR REMAINING LIFETIME PROGNOSIS

Steven M. Cox, BS, MS
Major, USAF

Approved: Date

Jeffrey P. Kharoufeh (Chairman)	
Dennis W. Quinn (Dean's Representative)	
Kenneth W. Bauer, Jr. (Member)	
Mark E. Oxley (Member)	

Accepted:

Robert A. Calico, Jr. Dean, Graduate School of Engineering and Management	Date

Table of Contents

Page

List of Figures vi

List of Tables vii

List of Abbreviations viii

Abstract ix

Acknowledgements x

I. Introduction 1
- 1.1 Background 1
- 1.2 Problem Definition and Methodology 5
- 1.3 Dissertation Outline 6

II. Literature Review 7
- 2.1 Systems Prognosis 7
- 2.2 State Estimation Approaches to Prognosis 10
  - 2.2.1 Regression Techniques 10
  - 2.2.2 Time Series Analysis 12
  - 2.2.3 Kalman Filtering 18
  - 2.2.4 Evolutionary Strategies 21
- 2.3 Stochastic Shock and Wear Models 24
- 2.4 Failure and Degradation-Based Reliability 33

III. Mathematical Model Description 37
- 3.1 Markovian Environment 38
- 3.2 Full Lifetime Distribution 42
  - 3.2.1 Time-Nonhomogeneous Markov Environment . 42
  - 3.2.2 Time-Homogeneous Markov Environment . . . 52
- 3.3 Moments of the Lifetime Distribution 55
  - 3.3.1 Matrix-Exponential (ME) Distributions 55
  - 3.3.2 ME Distribution Moments 56
- 3.4 Remaining Lifetime Distribution 62
- 3.5 Numerical Methods 63
  - 3.5.1 Observable Environment 64
  - 3.5.2 Observable Degradation 66
  - 3.5.3 Data Sources and Experimentation 74

Page

3.5.4 Goodness-of-Fit Test 75
3.5.5 Illustrative Examples 77

IV. Homogeneous Semi-Markov Environments 96
4.1 Review of Semi-Markov Processes 96
4.2 Lifetime Distribution Results 98
4.3 Illustrative Examples 106
4.3.1 Exponential State Holding Times 106
4.3.2 Hyper-exponential State Holding Times 108
4.3.3 Erlang State Holding Times 109

V. Phase-Type Approximations 112
5.1 Properties of PH-Distributions 112
5.2 Approximating a General Distribution 116
5.2.1 Summary of Methods 117
5.2.2 Minimal PH-Distribution 119
5.2.3 Choosing an Appropriate PH-Distribution . . 121
5.2.4 Comparison of Techniques 124
5.3 Conversion to a Markov Environment 139
5.3.1 Transformation Process 139
5.3.2 Lifetime Distribution Comparisons 141
5.3.3 Observable Degradation Comparisons 148

VI. Conclusions, Recommendations, and Future Work 156
6.1 Dissertation Contributions 156
6.2 Recommendations 157
6.3 Future Research 157

Bibliography 160

List of Figures

Figure		Page
2.1.	Artificial Neural Network	22
3.1.	Sample path of a continuous-time Markov chain.	39
3.2.	An example of a dendrogram using 2000 observations.	69
3.3.	Dendrogram with three clusters.	69
3.4.	Values for F_K showing 10 or 11 clusters.	72
3.5.	A sample of five linear degradation paths.	75
3.6.	Crack propagation for 68 samples of 2024-T3 aluminum alloy.	75
3.7.	A sample of five linear degradation paths.	83
3.8.	Piecewise approximations, $M = 10$, to the linear degradation paths.	83
3.9.	Simulated versus approximated degradation sample paths.	84
3.10.	Analytical CDF versus Simulated CDF	84
3.11.	CDF comparison with $t_{20} = 20$ (left) and CDF with $t_{50} = 20$ (right).	87
3.12.	CDF comparisons with $t_{100} = 20$ (left) and CDF with $t_{200} = 20$ (right).	87
3.13.	CDF comparison with $t_{500} = 20$.	88
3.14.	Distributions with known 5-state (left) and 10-state (right).	90
3.15.	Comparison of cumulative probability values.	93
4.1.	Sample path of a semi-Markov process.	97
5.1.	A three-phase general PH-distribution.	114
5.2.	A three-phase acyclic PH-distribution.	115
5.3.	A three-phase Coxian distribution.	115
5.4.	A three-phase Erlang distribution.	116
5.5.	Phase-type approximation of a Weibull(2.5, 3) cdf.	127
5.6.	Phase-type approximation of a Beta(2.0, 4.0) cdf.	128
5.7.	Phase-type approximation of a Gamma(2.5, 2.0) cdf.	129
5.8.	Phase-type approximation of a Beta(1.2, 4.0) cdf.	131
5.9.	Phase-type approximation of a Gamma(1.3, 2.0) cdf.	132
5.10.	Phase-type approximation of a Weibull(3.0, 1.03) cdf.	133
5.11.	Phase-type approximation of a Gamma(0.7, 2.0) cdf.	135
5.12.	Phase-type approximation of a Weibull(3.0, 0.6) cdf.	136
5.13.	Phase-type approximation of a Beta(0.1, 2.0) cdf.	137
5.14.	Graphical depiction of modified state transition diagram ($K = 3$)	140
5.15.	CDF comparisons for 3-state SMP, 12-state CTMC, and 13-state CTMC.	145
5.16.	CDF comparisons for 3-state SMP and 16-state CTMC environments.	147
5.17.	Lifetime comparisons of a system subject to a SMP environment.	154

List of Tables

Table		Page
3.1.	Cramér-von Mises test statistics for Case I (κ^*=0.461, $\alpha = 0.05$).	82
3.2.	Estimated generator matrices for each M at three times.	85
3.3.	Estimated degradation rates for each M at three times.	86
3.4.	Cramér-von Mises statistic for Models 1-4 (κ^*=0.461 at $\alpha = 0.05$).	88
3.5.	Values of $F_K(\times 10^5)$	89
3.6.	Cramér-von Mises statistic for Models 1-4 (κ^*=0.461 at $\alpha = 0.05$).	90
3.7.	Cramér-von Mises statistic for Model 1 and Model 3 (κ^*=0.461 at $\alpha = 0.05$).	91
3.8.	Mean lifetime and remaining lifetimes ($\times 10^5$), ξ_0 = actual $m^{(1)}(x)$.	92
3.9.	Values of $F_K(\times 10^4)$.	93
3.10.	Cramér-von Mises test statistic ($\kappa^* = 0.461$ at $\alpha = 0.05$).	94
5.1.	Lower moments and c^2 for various distributions.	125
5.2.	Weibull(2.5, 3.0) versus PH approximations.	126
5.3.	Beta(2.0, 4.0) versus PH approximations.	128
5.4.	Gamma(2.5, 2.0) versus PH approximations.	129
5.5.	Beta(1.2, 4.0) versus PH approximations.	130
5.6.	Gamma(1.3, 2.0) versus PH approximations.	132
5.7.	Weibull(3.0, 1.03) versus PH approximations.	134
5.8.	Gamma(0.7, 2.0) versus PH approximations.	135
5.9.	Weibull(3.0, 0.6) versus PH approximations.	136
5.10.	Beta(0.1, 2.0) versus PH approximations.	137
5.11.	Summary of PH-distribution comparisons.	138
5.12.	Cramér-von Mises test statistics: PH-distribution approximation techniques. (κ^*=0.461, $\alpha = 0.05$).	145
5.13.	Values of $F_K(\times 10^6)$	150
5.14.	Estimated degradation rates.	150
5.15.	Cramér-von Mises statistic for SMP conversion (κ^*=0.461, $\alpha = 0.05$).	155

List of Abbreviations

Abbreviation	
ADT	Accelerated Degradation Tests
AL	Autonomic Logistics
ALIS	Autonomic Logistics Information System
ALT	Accelerated Life Tests
ANOVA	Analysis of Variance
AR	Autoregressive
ARIMA	Autoregressive-Integrated Moving Average
CDF	Cumulative Distribution Function
CNC	Computerize Numeric Control
CTMC	Continuous-Time Markov Chain
DARTS	Diagnostic Analysis and Repair Tool Set
DTMC	Discrete-Time Markov Chain
EC	Erlang-Coxian
EDAPS	Embedded Diagnostics and Prognostics Synchronization
IFR	Increasing Failure Rate
IHRA	Increasing Hazard Rate Average
JSF	Joint Strike Fighter
KS	Kolmogorov-Smirnov
LEAP	Life Extension Analysis and Prognostics
LST	Laplace-Stieltjes Transform
MA	Moving Average
MAD	Maximum Absolute Deviation
ME	Matrix-Exponential
MMSE	Minimum Mean Squared Error
NICHD	National Institute of Child Health and Development
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
PH	Phase-type
PHM	Prognostics and Health Management
SIDS	Sudden Infant Death Syndrome
SME	Subject Matter Experts
SMP	Semi-Markov Process
TEDANN	Turbine Engine Diagnosis using Artificial Neural Networks
UAVs	Unmanned Aerial Vehicles

AFIT/DS/ENS/04-01

Abstract

This dissertation is concerned with the development of implementable analytical models for the estimation of the remaining lifetime probability distribution of a component subject to a randomly evolving environment. The models incorporate estimated parameters via environmental or degradation measures obtained from component sensors. We consider three distinct stochastic process models for the random environment: a temporally nonhomogeneous Markov environment, a temporally homogeneous Markov environment, and a temporally homogeneous semi-Markov environment. The hybrid approach unites real environment state or degradation measures with analytical, stochastic failure models to numerically compute the distributions and their moments. Additionally, it is shown that the lifetime distributions resulting from the homogeneous Markov environment and a special case of the nonhomogeneous Markov environment are distributions of the matrix-exponential type. Because the lifetime distribution in the semi-Markov case is computationally intensive, we instead utilize phase-type (PH) approximations that transform the semi-Markov environment to a time-homogeneous Markov environment. The numerical experiments indicate that the analytical techniques developed in this research hold great promise for remaining lifetime prognosis in a variety of contexts.

Acknowledgements

I was convinced that the greatest lessons learned over the last three years would be purely academic, but the lessons that my Lord and Savior Jesus Christ have taught me go well beyond academics. Learning to rely upon Him and to have a good attitude in spite of the circumstances are the ultimate contributions of this dissertation. I thank Christ above all others.

I thank my family who struggled with me every step of the way and I apologize for not learning to have a good attitude earlier. I cannot say enough about my amazing, wonderful wife. Taking care of our four kids and maintaining a great attitude herself was just incredible. I remarked to my neighbor one day, who happens to be the best neighbor I have encountered, that my wife is my hero. For my four terrific kids, you are all great and I thank you for your abilities to keep a smile on my face and to help me constantly realize there is something other than academics.

I also would not have completed this dissertation without the help of my advisor, Dr. Jeff Kharoufeh. His attention to detail and direction were tremendously helpful. He passed on some skills that I will value the remainder of my life. He is undoubtedly, one of the most brilliant men I have ever encountered. I also thank Dr. Ken Bauer and Dr. Mark Oxley for putting up with my many blunders.

Lastly, I want to thank the many people behind the scenes. The friendship and encouragement of those who followed us from USAFA helped pull us through on many occasions. I relied upon the many prayers from my small group at Faircreek Church, the Men's Prayer group at AFIT, my Emmaus share groups and so many others. All Praise to God!!!

Steven M. Cox

13 August 2004

HYBRID STOCHASTIC MODELS FOR REMAINING LIFETIME PROGNOSIS

I. Introduction

1.1 Background

The overarching mission of the United States Air Force is to "fly, fight, and win." Capabilities and technologies which improve the means by which the US Air Force flies, fights, and wins are continuously sought. For example, unmanned aerial vehicles (UAVs) have greatly increased the amount of time an aircraft remains in flight. Laser and satellite guided munitions have drastically improved the ability of the US Air Force to successfully complete their missions. Stealth aircraft have tremendously enhanced the ability to win with their capacity to attack enemy targets virtually unseen. The Joint Strike Fighter (JSF), the newest fighter aircraft in production, is part of an ongoing effort to address all three aspects of the Air Force mission. Its state-of-the-art munitions delivery technology seeks to provide unparalleled capability to put "bombs on target" and destroy enemy aircraft. Its most up-to-date stealth technology seeks to conceal its very presence from enemy radars. These technology enhancements, and the overall functionality of the aircraft, rely upon its operational status. If the aircraft is down for repairs or waiting for parts, the "fly and fight" portions of the mission are obviously impacted while the "win" portion may well be impacted by how long it takes to return the aircraft to an operational status. Thus, maintenance and logistics teams continually seek methods to return aircraft to an operational status in an efficient manner.

An area which will provide a great enhancement in the JSF's ability to remain airborne rests in maintenance and logistics. Current procedures call for an aircraft

to take off, fly its mission, return to base, land, and then be readied for the next mission. In this protocol, maintenance and (to a lesser degree) logistics activities, do not begin until after the aircraft has landed. Maintenance personnel obtain a list of malfunctioning systems from both the pilot and aircraft diagnostics and then begin the process of locating and repairing (or replacing) failed components. The amount of time required to locate failures, coupled with the availability of replacement parts, can severely degrade the operational effectiveness of the aircraft. To lessen this impact, a JSF product team [35] is working towards the development of an autonomic logistics (AL) program which seeks to produce high sortie generation rates, high mission reliability rates, a small logistics footprint, and a significant reduction in ownership costs including reduced manpower.

The AL program relies on four key elements: a prognostics and health management (PHM) system, an autonomic logistics information system (ALIS), a maintainer, and a support system. The PHM system and ALIS begin functioning as soon as the aircraft is operational. In theory, the PHM system examines the aircraft's critical components, detects any faults or failures and provides future failure probabilities on currently operating components. This information is transmitted to the maintainer via the ALIS. From an operational point of view, the value of such a system is obvious because the maintainer receives advance notice of an impending failure via the ALIS and is able to initiate the maintenance and logistic processes to prepare for any repairs or replacements *before* the aircraft lands. However, the technology to fully implement the AL program has not been completely developed. This dissertation focuses on a portion of the technology to implement the PHM system of the AL program.

The PHM system includes both diagnosis and prognosis of the system and its associated components. The science of diagnostics can be defined as the detection, isolation, and identification of a fault or failure whereas prognosis attempts to estimate, with some degree of certainty, the *remaining* useful lifetime of the system. In

diagnosis, a fault indicates degraded operation whereas failure indicates operation has ceased. Routine checkups, instantaneous failures, degraded performance, and other criteria may indicate a fault or failure has occurred. Once detected, methods to isolate the impacted components and assess the level of damage to those components are employed. If not fully isolated and identified, additional faults and failures related to the original failure are more likely. Prognosis naturally relies upon, but is distinct from the diagnosis of a system. The ultimate goal of prognosis is to assess the current health of a system in order to make inferences about the future health of the system. The ability to predict when a component will fail has obvious benefits. These benefits include reduced risk of catastrophic failure, utilization of the full lifetime, a capability to plan for a failure instead of reacting to a failure and secondary damage, as well as many other benefits. If a failure is accurately diagnosed, then prognosis obviously has no value. However, if diagnosis indicates a system is operating with or without faults, then prognosis is useful to assess the remaining lifetime of critical system components.

While the value of remaining lifetime prognosis is clear, implementation is nontrivial. Diagnosis has been used for many years in mechanical systems and components, and the methods to automatically diagnose these systems correctly have improved greatly by enhancements in detection, isolation and identification of faults and failures. However, implementing a prognosis capability in such systems has not been realized to the same degree. This is partly due to the increased reliability of components and systems wherein it may take years to observe a single fault or failure (e.g. seat belts in a car). In such cases, there is little, if any, historical lifetime data on components, thus complicating the task of lifetime estimation. If sufficient lifetime data are available, lifetime distributions (parametric or nonparametric) can be developed to predict a component's lifetime using standard statistical techniques. However, without lifetime data, some have resorted to periodic replacement policies, but this type of maintenance strategy may not fully utilize the useful life of the

component. Still others use an event-driven replacement strategy where components are replaced after the occurrence of a fixed number of events, such as the number of aircraft landings. Optimal maintenance strategies continue to be a major concern for analysts in military and commercial settings. Therefore, a prognosis methodology, not contingent on lifetime observations, will ultimately enable the PHM system implementation.

Smith, *et al.* [81] determined that the prognostic capability of the PHM system supporting the AL program is crucial to the development of the Joint Strike Fighter. Henley, *et al.* [31] have concluded that "AL will not work without PHM." Since system diagnosis is well defined and currently implemented on many systems, successful AL hinges upon successful prognosis. The major contribution of this dissertation is the development of a hybrid approach that combines analytical techniques with real-time component degradation data to determine the probability distribution of the full and remaining lifetime of components for the purpose of system prognosis and possible implementation into the PHM system. The uniqueness of this contribution rests in the specification of the lifetime distribution using only degradation data in lieu of failure time observations. Current methodologies (e.g. time series analysis and artificial neural networks) estimate lifetimes and prediction intervals associated with those lifetimes. These techniques, however, require sufficient data to observe trends or train and validate the resulting model for each component. The techniques developed in this dissertation will be sufficiently generic to provide the lifetime distribution for degradation data provided by sensors attached to a component. The proposed techniques require knowledge of the random environment in which the system operates. Additionally, while the results are highly relevant to Air Force applications, a broader contribution will be made to the growing field of degradation-based reliability analysis.

This dissertation does not seek to improve diagnosis capabilities, but is focused on providing contributions in the area of systems prognosis. Arguably, the most

important hurdle in the implementation of systems prognosis is the ability to assess the future health of a component given no failure time observations. We provide a technique to perform systems prognosis without failure time observations via the determination of the remaining lifetime distribution using degradation data and associated information from the random environment in which the system operates.

1.2 Problem Definition and Methodology

This dissertation is primarily focused on the development of analytical models that can be employed to estimate the remaining lifetime probability distribution by using degradation measures obtained from existing component sensors. These distribution estimates are extremely important in the overall context of systems prognosis to assess the current health of the component and predict its remaining lifetime.

The approach of this dissertation combines analytical models found in wear processes with degradation-based reliability to provide a methodology to obtain the remaining lifetime distribution of a single-unit system. While lifetime observations are rarely available, sensors attached to a component are capable of providing information directly indicating the degradation of the component and/or current environmental conditions. However, given only real-time sensor measures, there does not exist a method to analytically compute the remaining lifetime distribution of the component.

This dissertation develops analytical results for the lifetime distribution that incorporate real data, provided by sensors, of a single-unit system operating in a random environment modelled as a nonhomogeneous and homogeneous, continuous-time Markov chain. We consider solving this problem under two scenarios: when the environment is observable and when degradation is observable. We show through goodness-of-fit tests that this estimated lifetime distribution is statistically equiva-

lent to simulated lifetime distributions. Additionally, it is shown that the remaining lifetime distribution is determined via the lifetime distribution resulting in the capability to perform systems prognosis. These results are then extended to account for a more general environment modelled as a semi-Markov process. Finally, it will be shown that this process requires phase-type (PH) distribution approximations to numerically implement the analytical lifetime distribution.

1.3 Dissertation Outline

Chapter II examines current prognosis procedures in military and industrial systems as well as other prediction techniques. In Chapter III, we derive analytical results for the lifetime distribution by consideration of the nonhomogeneous and homogeneous Markovian environment in which the unit operates. We then show these distributions are a subset of a much larger class of distributions whose moments are easily obtainable. Chapter III also provides and demonstrates the procedures for numerical implementation of these analytical results.

Chapter IV examines a more general random environment; however, our main results in this chapter are extremely difficult to implement numerically. This difficulty is illustrated through three examples that result in complex matrix equations. Working around this complexity provides the basis for Chapter V wherein we apply techniques that approximate the more general environment and allow us to numerically evaluate the lifetime distribution using the techniques developed in Chapter III. Finally, we highlight the contributions and recommendations of this dissertation and suggest future areas of research in Chapter VI.

II. Literature Review

The overall aim of this research is to obtain the remaining lifetime distribution for the purpose of performing systems prognosis. Some prognosis methods already exist, as shown in Section 2.1, but do not necessarily incorporate degradation data or provide the remaining lifetime distribution. The methods given in Section 2.2 incorporate a steady stream of empirical data which could include degradation data. These are primarily state estimation techniques with the objective to predict some future state given the history of the process. Thus, their intention is not to provide the remaining lifetime distribution. Many analytical models (e.g. shock and wear models) provide a closed-form solution for the lifetime distribution, but provide little guidance on numerical implementation and may not be able to incorporate real-time degradation data. Lastly, Section 2.4 compares and contrasts failure-based approaches, which require failure time observations, and degradation-based approaches which require simplifying assumptions to estimate the lifetime and remaining lifetime distributions.

2.1 Systems Prognosis

This section provides a sampling of some organizations that have incorporated systems prognosis in their operations. Though the majority of this discussion is devoted to military applications, it does include industrial and medical applications as well.

The US Army [22] designated that material developers should emphasize diagnosis and prognosis in the design, development and improvement of their equipment. Hence, the US Army has attempted to incorporate various aspects of systems prognosis. Kangas, *et al.* [38] addressed the health monitoring of the M1A1/A2 Battle Tank. The turbine engine diagnosis using artificial neural networks (TEDANN) methodology focused on diagnostics, but did provide a capability to predict future

availability of the vehicle and schedule maintenance via trend analysis. Greitzer, *et al.* [29] indicated that the short-term and long-term trend analysis capability in TEDANN is based on linear regression. Additionally, the author introduced a methodology called life extension analysis and prognostics (LEAP) and applied it to the US Army's M1 Abrams tank [28] and a diesel locomotive engine [94]. Su, *et al.* [82–84], were funded by the US Army Logistics Integration Agency to develop a prognostics framework focused on an entire system rather than a single component of that system. This framework was initially based upon the diagnostic analysis and repair tool set (DARTS) and has the ability to accept prediction information from neural networks, time/stress measurement devices, vibration monitoring, oil monitoring, sensors, trend analysis, and statistical analysis. Parts reliability, degradation data, failure mode data, and other prognostic related data were also integrated into the framework. Furthermore, Su, *et al.* [83] provided a summary of over sixty articles surveying prognostics research in the military, academia, and industry. His summary included applications such as power plants, US Navy ships, wind tunnels, and helicopters using techniques such as artificial and polynomial neural networks, Kalman filters, rule based expert systems, fuzzy logic, and statistical network modeling. Su, *et al.* [83] indicated these were prognostic techniques, but none provided the remaining lifetime distribution.

In spite of these attempts to incorporate systems prognosis in the US Army, Keeney, *et al.* [43] mentioned that embedding diagnosis and prognosis capabilities is a major challenge and examined numerous questions associated with this task. He indicated that the process was not complete due to funding limitations and provided requirement specifications that have derailed systems diagnosis and prognosis integration. However, to speed the process, Keeney, *et al.* [43] indicated that the US Army Logistics Integration Agency had initiated a new project called the Embedded Diagnostics and Prognostics Synchronization (EDAPS) with the purpose of pulling together all current US Army efforts in embedded diagnostics and prognostics.

The US Navy has also adopted the previously mentioned prognostics framework [66], but it appears to be mainly a diagnostic framework with the ability to incorporate systems prognosis information on individual components as that information becomes available. Hardman, *et al.* [30] helped generate a US Navy strategy to develop diagnosis and prognosis procedures for helicopter drivetrains using rule-based and model-based analysis techniques. The authors indicated that extrapolation of key data, statistical parameters of diagnostic indices, and trend analysis are techniques that enable failure prediction. However, it was stated that each failure mode for a specific component type might require a separate analysis.

In the private sector, manufacturers have also adopted the concept of systems prognosis. Caterpillar [13] attached pump vibration sensors to hydraulic pumps on backhoes for the purpose of detecting a 10% flow loss from an established baseline. They trained and tested a neural network on autoregressed, filtered data. BorgWarner Cooling Systems [11] published a brochure on their fan drives for light trucks and sport utility vehicles claiming the fan speed feedback signal allows for diagnosis and prognosis of the electronic control module. Cummins Inc. [20] mention in the industrial specifications for their QST30 engines that the CELECT^{TM} electronic engine management provides electronics for diagnosis and prognosis. There are also purchasable options on the engine that provide trend data for assistance in maintenance decisions.

In the medical community, prognosis is mainly associated with the determination of the most important factors related to the detection, progression, and prevention of illnesses. For example, the National Institute of Child Health and Development (NICHD) [63] published a pamphlet on sudden infant death syndrome (SIDS) and stated that improvements in screening tools relied upon knowing the abnormalities that cause death where proper prognosis and diagnosis would play a major role in these tools. Park, *et al.* [71] examined a new cancer therapy by noticing how the inhibitor p27 suppressed the growth of established lung cancer. Previous

research in tumors had indicated that p27 possibly had a prognosis capability. Most all medical applications attempt to find the most important factors behind certain life shortening diseases and ailments. Once these factors are known, procedures may be applied to predict the remaining lifetime of a patient.

Many of the prognosis techniques mentioned in this section rely upon state estimation techniques which are examined in the next section. We notice that none of these current system prognosis techniques allow for the incorporation of degradation data, and nor do they provide the remaining lifetime distribution.

2.2 State Estimation Approaches to Prognosis

If a sufficient amount of time-dependent data is available, then stochastic state estimation techniques are often used to determine the state, or functional value of the system, at a future point in time. Most of the applications in Section 2.1 incorporate these techniques that include regression, time series, Kalman filters, neural networks, and others. Probability distributions are not usually estimated from state estimation techniques, but can be if sufficient data exists. These techniques rely heavily on past data to predict future performance. In some cases where multiple data sets exist, it is feasible to use statistical techniques or simulation which can provide an estimate of the failure time distribution. In the following subsections, a few common state estimation procedures are reviewed.

2.2.1 Regression Techniques

This section provides an overview of the concepts and techniques associated with regression analysis. Regression analysis uses the existing data and determines the relationships, if any, between the measurable outcome and the variables contributing to that outcome (e.g. life expectancy is the outcome and exercise and diet are the variables contributing to that outcome). Neter, *et al.* [64] provided the

framework on the statistical relation for the purpose of prediction. A general linear regression model is given by

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \qquad i = 1, \ldots, n \tag{2.1}$$

where Y_i is a random variable denoting the value of the i^{th} trial's response, β_0, β_1, $\ldots, \beta_{p-1}$ are estimated parameters, $X_{i,1}$, $X_{i,2}$, $\ldots$, $X_{i,p-1}$ are the values of the predictor, or contributing variables, and ε_i is the random error with mean = 0, variance = σ^2, and covariance = 0. The term linear results from the estimated parameters, not the predictor variables. For example, Equation (2.2) is a linear model whereas Equation (2.3) is a nonlinear model.

$$Y_i = \beta_0 + \beta_1 X_{i,1}^2 + \beta_2 \sqrt[3]{X_{i,2}} + \beta_3 X_{i,1} X_{i,2}^2 + \varepsilon_i. \tag{2.2}$$

$$Y_i = \beta_0 + \beta_1^2 X_{i,1}^2 + \beta_2 \sqrt[3]{X_{i,2}} + \beta_3 X_{i,1} X_{i,2}^2 + \varepsilon_i. \tag{2.3}$$

Define the regression function for the regression model in Equation (2.1) as

$$\mathbb{E}[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1}. \tag{2.4}$$

Regression analysis seeks to estimate the parameters of the regression function, $\beta_0, \beta_1, \ldots, \beta_{p-1}$, in order to find a representative model by the method of least squares. The method of least squares defines a variable Q, where

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i,1} - \beta_2 X_{i,2} - \ldots - \beta_{p-1} X_{i,p-1})^2 \tag{2.5}$$

and attempts to find estimates for β_0, β_1, $\ldots$, β_{p-1}, denoted by b_0, b_1, $\ldots$, b_{p-1}, which minimize Q for the observations (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n). The simultaneous solution to the equations formed by taking the derivative of Q with respect to β_0,

$\beta_1, \ldots, \beta_{p-1}$ provides the least squares estimates, b_0, b_1, ..., b_{p-1}. Least squares estimates are desired because they are unbiased and have minimum variance resulting in

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_{p-1} X_{p-1}. \tag{2.6}$$

The method of maximum likelihood can also be used to estimate β_0, β_1, ..., β_{p-1} if the probability distribution of the error terms is known.

Li, *et al.* [55] discussed mainly diagnosis results based upon a regression model, but also included some details on prognosis. They examined an adaptive prognostics approach where a future bearing defect size was calculated at time $t+\Delta$ ($\Delta > 0$) given the bearing running condition and defect size at time t. This adaptive algorithm, based on a recursive least squares algorithm applied to a defect power law-based propagation model, was then employed to account for the time-varying behavior and used to predict future impending failures. Additionally, as mentioned in Section 2.1, regression is employed for prognosis in many medical applications. Typical response variables include life expectancy or recovery time with predictor variables such as age, family history, personal habits, and amount of time spent exercising in a week.

The details in this section scarcely scratch the surface of regression analysis. Neter, *et al.* [64] provide much more detail on linear and nonlinear regression models as well as the dangers of extrapolating beyond the observed data in their text. However, regression is not the only state estimation method used for prognosis.

2.2.2 Time Series Analysis

In addition to regression analysis, time series or autoregressive-integrated moving average (ARIMA) processes, also known as trend analysis, is a common state estimation technique used for prognosis. The information in this section can be found

in Box and Jenkins [12] unless specifically cited otherwise. The following notation is helpful for the methods presented in this section.

μ is the mean of the time series data
$z_1, z_2, \ldots, z_n$ are time ordered observations
$w_1, w_2, \ldots, w_n$ are differenced time ordered observations
B is the backshift operator where $Bz_t = z_{t-1}$ and $B^m z_t = z_{t-m}$
I is the identity operator
∇ is the difference operator where $\nabla z_t = z_t - z_{t-1} = (I - B)z_t$
$a_t, a_{t-1}, a_{t-2}, \ldots \sim N(0, \sigma_a^2)$ are known as a white noise process
σ_a^2 is the variance of the white noise process a_t
$\phi_1, \phi_2, \ldots, \phi_p$ are the unknown parameters of an autoregressive process
$\theta_1, \theta_2, \ldots, \theta_q$ are the unknown parameters of a moving average process
$\phi(B) = I - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is an operator of order p
$\theta(B) = I - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is an operator of order q

An ARIMA model is a generic construct which incorporates autoregressive (AR) processes, moving average (MA) processes, and a capability to account for non-stationary data. Given $\tilde{z}_t = z_t - \mu$, an AR process of order p is mathematically defined as

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t \tag{2.7}$$

which is similar to Equation (2.1) and can be rewritten as

$$\phi(B)\tilde{z}_t = a_t. \tag{2.8}$$

Observed data provides estimates for $\mu, \phi_1, \phi_2, \ldots, \phi_p$. An MA process of order q is defined as

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{2.9}$$

and can be rewritten as

$$\tilde{z}_t = \theta(B)a_t \tag{2.10}$$

where observed data provides estimates for $\mu, \theta_1, \theta_2, \ldots, \theta_q$.

A non-stationary model must be transformed into a stationary model before the AR and/or MA techniques are appropriate. This transformation normally occurs through differencing, $(I-B)^d$, but can also be accomplished by taking the logarithm of the time series data. Therefore, a complete ARIMA model of order (p, d, q) is mathematically defined as

$$\phi(B)(I - B)^d \tilde{z}_t = \theta(B) a_t. \tag{2.11}$$

This model can describe both stationary and non-stationary time series but requires a significant amount of data to estimate μ, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$.

Before any parameters can be estimated, the order for each process in the ARIMA model must be determined. The purpose of this step is to make the process stationary, if need be, by determining d, and then determine the number of parameters to be estimated by identifying both p and q. Box and Jenkins [12] specifically state that the methods to identify the order of an ARIMA model are inexact and that "statistically 'inefficient' methods must necessarily be used." Thus, graphical methods are very helpful in the determination of these orders. Personal judgement used to determine the values of p, d, and q are often based on the estimated autocorrelation and partial autocorrelation functions. Given N is the total number of observations, $\bar{z}$ is the mean of these observations, and $K \leq \frac{N}{4}$, the kth lag autocorrelation ρ_k is estimated by

$$r_k = \frac{c_k}{c_0} \tag{2.12}$$

where

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \qquad k = 0, 1, 2, \ldots, K. \tag{2.13}$$

The autocorrelation function is a plot of the values of r_k at each lag k. Given the autocorrelation function, if r_k approaches zero, or 'dies out', quickly, the time series is considered stationary. If the autocorrelation does not approach zero quickly, then differencing is required. The value for d is determined using this autocorrelation function where d, normally 0, 1, or 2, is increased by one until the series is deemed stationary. With d estimated, the autocorrelation and partial autocorrelation functions are used to find p and q. Using r_k as estimates for ρ_k, the following system of equations are solved for $\hat{\phi}_{kk}$,

$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}. \tag{2.14}$$

The partial autocorrelation function is formed by plotting $\hat{\phi}_{kk}$ at each lag k and the order p of an autoregressive process is equal to, in most cases, the last 'nonzero' lag p of the partial autocorrelation function. Additionally, the order q of a moving average process is equal to, in most cases, the last 'nonzero' lag q of the autocorrelation function. The term 'nonzero' defines a cutoff point where the remaining $k-q$ and $k-p$ lags in the autocorrelation and partial autocorrelation functions, respectively, come close to zero. If there is no cutoff point in either the autocorrelation or partial autocorrelation functions, then a combined ARMA process is required. In this combined scenario, either the autocorrelation function or partial autocorrelation function will be infinite with damped exponentials and/or sine waves after the first $q-p$ or $p-q$ lags, respectively. It should be clear that there is some leeway in the identification of p, d, and q. Once all three orders are determined, the ARIMA model parameters can be estimated.

An ARIMA(p, d, q) with $d \neq 0$ will have $n = N - d$ observations, $w_1, w_2, \ldots, w_n$ and can be written as

$$a_t = \tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \phi_2 \tilde{w}_{t-2} - \ldots - \phi_p \tilde{w}_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q} \quad (2.15)$$

where $\mathbb{E}[w_t] = \mu$ and $\tilde{w}_t = w_t - \mu$. Chatfield [17] shows that, when estimating the parameters of an ARIMA$(p, d, 0)$ model, the parameters, $\phi_1, \phi_2, \ldots, \phi_p$, can be estimated with least squares by minimizing

$$S = \sum_{t=p+1}^{N-d} [\tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \phi_2 \tilde{w}_{t-2} - \ldots - \phi_p \tilde{w}_{t-p}]^2 \quad (2.16)$$

with respect to $\phi_1, \phi_2, \ldots, \phi_p$. Additionally, Box and Jenkins [12] and Chatfield [17] show that for N reasonably large, the solution to the matrix equation $\mathbf{R}\hat{\boldsymbol{\phi}} = \mathbf{r}$ provides estimates close to the least squares estimates in Equation (2.16) where

$$\mathbf{R} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & 1 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix} \quad (2.17)$$

$$\hat{\boldsymbol{\phi}}^T = \begin{bmatrix} \hat{\phi}_1 & \hat{\phi}_2 & \ldots & \hat{\phi}_p \end{bmatrix} \quad (2.18)$$

$$\mathbf{r}^T = \begin{bmatrix} r_1 & r_2 & \ldots & r_p \end{bmatrix}. \quad (2.19)$$

Chatfield [17] mentions that "efficient explicit estimators" for an ARIMA$(0, d, q)$ process cannot be found. Therefore, an iterative technique must be used with suitable starting values for $\theta_1, \theta_2, \ldots, \theta_q$. While Chatfield [17] provided a starting point for θ_1, Box and Jenkins [12] provided a system of equations where initial estimates

for all θ's are found by simultaneously solving

$$r_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} \tag{2.20}$$

for $k = 1, 2, \ldots, q$. Given initial estimates and the portion of Equation (2.15) pertaining to moving averages, Chatfield [17] provided the recursive equations,

$$\begin{aligned} a_1 &= \tilde{w}_1 \\ a_2 &= \tilde{w}_2 - \theta_1 a_1 \\ a_3 &= \tilde{w}_3 - \theta_1 a_2 - \theta_2 a_1 \\ &\vdots \\ a_n &= \tilde{w}_n - \theta_1 a_{n-1} - \theta_2 a_{n-2} - \cdots - \theta_q a_{n-q} \end{aligned} \tag{2.21}$$

which are then used to calculate the sum of squares, $\sum_{t=1}^{n} a_t^2$. For $q \leq 3$, this method is repeated for updated estimates of θ_1, θ_2, and θ_3. The MA model, Equation (2.10), is then plotted with each set of estimates to find the least squares estimate which happens to be the lowest plot. If the order of the MA process exceeds three, Chatfield [17] discussed other iterative optimization techniques, which are not discussed here, that could be used. If additional data becomes available, the entire ARIMA process should be repeated. If the order of the ARIMA model has not changed, it is still necessary to re-estimate the parameters with additional data.

Jardim-Goncalves, *et al.* [34] used ARIMA models to predict when computerized numeric control (CNC) lathe and mill machines would fail. These machines were monitored with sound, vibration, and power consumption sensors in real time and the authors were able to forecast whether the machines required maintenance in future time periods given acceptable ranges on the monitored parameters. Patankar and Ray [72] examined the fatigue crack growth prediction problem with a forecasting model under variable-amplitude loading. The developed forecasting model was

shown to be adequate for real time applications. As seen, time series analysis, as presented here, can predict a future state at a future time. Another method that can perform this function is Kalman filtering.

2.2.3 Kalman Filtering

This section examines another state estimation technique called a Kalman filter with scalar states and scalar observations. A Kalman filter incorporates the signal embedded with noise and forms what can be considered a sequential minimum mean square error estimator (MMSE) of the signal. The signal is considered a state model. Kay [42] called the Kalman filter an important generalization of the Wiener filter because of its ability to handle non-stationary vector signals and noise as compared to the stationary scalar signals and noises required for the Wiener filter.

Define s_n, $n = 0, 1, \ldots, N-1$, as the signal to be estimated at a discrete time point n. The recursive scalar state equation, also called a first-order Gauss-Markov process, is

$$s_n = as_{n-1} + u_n, \quad n \geq 0 \tag{2.22}$$

where $|a| < 1$ is a weighting coefficient and u_n is white Gaussian noise with zero mean and variance σ_u^2. At $n = 0$, s_{-1} is the initial state where $s_{-1} \sim \mathbb{N}(\mu_s, \sigma_s^2)$ and μ_s and σ_s^2 are determined by the previous data. Additionally, s_{-1} is independent of all u_n. The scalar observation equation is defined as

$$x_n = s_n + w_n, \quad n \geq 0 \tag{2.23}$$

where each x_n is an observation and w_n is zero mean Gaussian noise with variance σ_n^2 which can change over time. Given observations, $\{x_0, x_1, \ldots, x_n\}$, filtering occurs when these observations are used to estimate s_n as n increases. The Kalman filter is the MMSE which recursively calculates the estimator $\hat{s}_n$. The estimator of s_n using

n observations, denoted by $\hat{s}_{n|n}$, is

$$\begin{aligned} \hat{s}_{n|n} &= \mathbb{E}[s_n|x_0, x_1, \ldots, x_n] \\ &= \hat{s}_{n|n-1} + \mathbb{E}[s_n|x_n] \end{aligned} \tag{2.24}$$

if x_n is uncorrelated with all previous observations. Since x_n is generally correlated with previous observations, a correction to the old estimator, $\hat{s}_{n|n-1}$ is required.

Kay [42] mentioned there is a portion of x_n which is uncorrelated with the previous observations and defines this portion as

$$\tilde{x}_n = x_n - \hat{x}_{n|n-1} \tag{2.25}$$

where $\hat{x}_{n|n-1} = \sum_{k=0}^{n-1} a_k x_k$ and a_k are optimal weighting coefficients. Equation (2.24) can now be rewritten with the correction as

$$\hat{s}_{n|n} = \hat{s}_{n|n-1} + \mathbb{E}[s_n|\tilde{x}_n]. \tag{2.26}$$

Since $\hat{s}_{n|n-1}$ is the prediction of s_n from $\{x_0, x_1, \ldots, x_{n-1}\}$, Equation (2.22) becomes

$$\hat{s}_{n|n-1} = a\hat{s}_{n-1|n-1} \tag{2.27}$$

because $\mathbb{E}[u_n|x_0, x_1, \ldots, x_{n-1}] = \mathbb{E}[u_n] = 0$.

Additionally, Kay [42] provided the derivations for the minimum prediction mean square error, Kalman gain, the correction, and the minimum mean square

error. Those equations are respectively,

$$
\begin{aligned}
M_{n|n-1} &= a^2 M_{n-1|n-1} + \sigma_u^2 && (2.28)\\
K_n &= \frac{M_{n|n-1}}{\sigma_n^2 + M_{n|n-1}} && (2.29)\\
\hat{s}_{n|n} &= \hat{s}_{n|n-1} + K_n(x_n - \hat{s}_{n|n-1}) && (2.30)\\
M_{n|n} &= (1 - K_n)M_{n|n-1}. && (2.31)
\end{aligned}
$$

This scalar Kalman filter formulation can be generalized to vector states with scalar observations and also to vector states and vector observations. These generalizations, however, are invalid given a nonlinear state equation and/or observation equation. The extended Kalman filter was developed to account for these nonlinearities.

Ray and Tangirala [75] used Kalman filters for the real time computation of fatigue crack dynamics as an alternative to solving the Kolmogorov forward equation. In a later paper, these authors [76] also examined fatigue crack growth prediction using Gauss-Markov processes which did not require solution of the extended Kalman filter equation. However, validation of the model with real data was limited in this scenario. Additionally, Ray and Tangirala [76] assumed that the crack length itself was distributed lognormally and combined state estimation procedures and parametric failure time distributions.

As seen in the title of the article [75], Kalman filtering is considered a prognosis technique by estimating some state value at a future point in time. Evolutionary strategies also claim to be a prognosis technique by providing information on the state of the system at a future point in time.

2.2.4 Evolutionary Strategies

Artificial neural networks, genetic algorithms, fuzzy logic, and other learning techniques comprise a class of approaches known as evolutionary strategies. These techniques have some ability to "learn" using past history and subsequently attempt to predict the state or outcome given a new set of input data. Hence, these techniques are the most frequently used in current prognosis procedures. Neural networks are examined in this section.

According to Tsoukalas and Uhrig [88], artificial neural networks work in a manner similar to actual neurons found in the human brain. Each neuron has dendrites which are the input paths, a soma which processes the inputs and an axon which is the output path. Neurons receive and transmit information to other neurons through synapses, or gaps between the dendrites and axons with chemicals that alter the flow of electrical charges. The artificial neural network has inputs or observations, $x_0, x_1, \ldots, x_n$, that act as dendrites. The inputs are pre-multiplied by a separate weight, $w_{i,j}$ representing the synapses. The soma in the artificial neural network performs two functions. It sums all the weighted inputs, $S_j = \sum_{i=1}^{n} w_{i,j} x_i$, for the neuron, j, and then transforms this summation for output. Tsoukalas and Uhrig [88] indicate that a transfer function acts as a nonlinear filter on each S_j which provides the output that travels along the axon for each neuron. The transfer function can take the form of a threshold function, a signum function, a sigmoidal function or other type of function and limits the range of the output. The most common activation function is the logistic function where

$$\Phi(S_j) = \frac{1}{1 + e^{-\alpha S_j}}. \tag{2.32}$$

An artificial neural network is formed by a collection of these artificial neurons. This network normally has an input layer, one or more hidden layers, and an output layer. Each layer, with the exception of the input layer, consists of a number of

neurons. Weights are designated for each input from the input layer to neurons in the hidden layer. Given the transfer function results from the hidden layer, another set of weights is applied as inputs to the output layer. The final transfer function results from the output layer form the overall outputs of the artificial neural network. An artificial neural network [88] with three inputs, four neurons in the hidden layer, and two neurons in the output layer is shown in Figure 2.1. While weights are attached to each arrow, only weights for input x_1 and neuron 7 are shown.

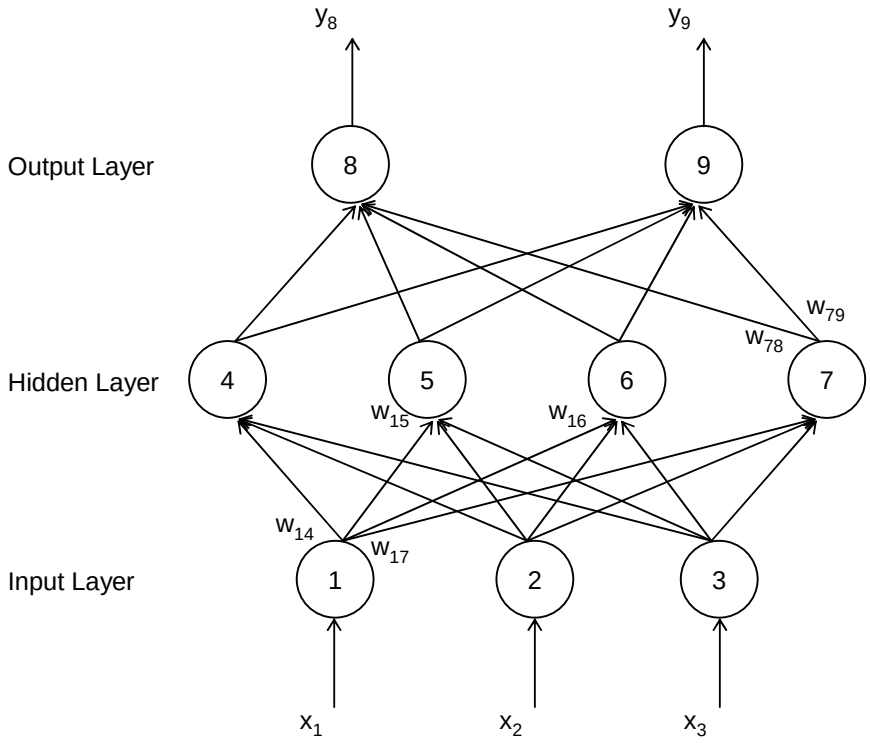

Figure 2.1: Artificial Neural Network

The application of artificial neural networks relies upon many sets of data and their associated outputs. The concept is to learn the patterns of association between the data inputs and their associated outputs for the purpose of prediction given another set of inputs. Initial weights in the network may assume any value, but Tsoukalas and Uhrig [88] indicate that smaller randomized weights work best from their experience. The network adjusts these weights by minimizing the error,

normally in the least squares sense, between the predicted and actual output. The process of adjusting the weights in neural networks is called *training* and occurs iteratively. Once training of the weights is complete, additional data can be used to validate or fine tune the model. Once fully trained, prediction using the neural network model can commence.

Li and Ray [56] examined the utility of using back-propagation neural networks to predict fatigue damage. Their methodology was used to decrease the computational time required for the conventional method of numerically solving nonlinear differential equations. In addition, the authors stated that this neural network could possibly be used for real time analysis of fatigue damage models and other types of failure models in addition to predicting remaining service life.

Other state estimation techniques exist. For example, Ray and Patankar [78] used a deterministic state-space model with state variables for the crack length and the crack opening stress in their fatigue crack propagation model. The model was modified by using Karhunen-Loève expansion of the crack length and a random process for the crack opening stress. The authors obtained a closed-form solution of the stochastic differential equation for the non-stationary statistic of the crack growth process. In a number of papers, Ray and his co-authors [56,72,75,76,78] used various state estimation techniques to model fatigue crack growth. The authors mentioned in these (and numerous other papers not referenced here) that these models could be used to predict the remaining life. In [77], Ray used his model for remaining life prediction and risk analysis examining both Virkler data [90] and Ghonem and Dore data [26] by generating the non-stationary probability distribution function of fatigue crack damage. Ray [77], assumed that both sets of data were lognormal distributed and used his model of fatigue crack growth to determine the parameters of the lognormal distribution for these data sets. Given the parameters, the author used Monte Carlo simulation to find the probability distribution function for the remaining lifetime. With respect to the Virkler data [90], Ray [77] compared his

generated probability distribution with empirical degradation levels at 11 mm, 14 mm, and 20 mm with good results.

These state estimation techniques, with respect to systems prognosis, are primarily used to predict the time when a given process will enter a failure state. Given multiple degradation paths, state estimation techniques can also, in some cases, estimate the failure distribution with current statistical techniques and some assumptions. Thus, the capability to determine the remaining lifetime distribution is quite limited with respect to state estimation techniques when relying upon the degradation data alone. We next examine analytical models that provide the close-form analytical solution of the lifetime distribution.

2.3 Stochastic Shock and Wear Models

The primary focus of this dissertation is the specification of the remaining lifetime distribution for a degrading component assuming that a monotonic failure path of degradation data is provided. Section 2.2 examined techniques focused mainly on state estimation real data. This section reviews analytical models that provide analytical results for lifetime distributions.

A discrete-time, finite-state shock model can be employed for the purpose of modelling cumulative damage to an individual component. In this basic form, such models provide a means to compute the cumulative distribution function of the random time required to reach a failure state. The failure state in the shock model corresponds to a prespecified level of cumulative damage which is assumed to be a monotonically increasing function of time. We formalize a shock model as follows. Define the random variable X_n as the damage state at time (or duty cycle) n with sample space $\Omega = \{1, 2, \ldots, b\}$ where b is an absorbing state. The stochastic process, $\{X_n : n \geq 0\}$, either remains in its current damage state or transitions to a higher damage state at each time step. Component failure occurs when the stochastic

process first enters the state b. There is no reliance on the history of the process which implies that $\{X_n : n \geq 0\}$ is a discrete-time Markov chain (DTMC) [50] satisfying the Markov property

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}. \quad (2.33)$$

Bogdanoff and Kozin [10] examine both a one-step and a multi-step shock model. In a one-step shock model, $P\{X_{n+1} = i | X_n = i\} = p_i$, $0 < p_i < 1$, meaning that in one time epoch, the system retains its current level of cumulative damage and does not transition to the next damage level. The cumulative damage increases to the next damage state with probability $P\{X_{n+1} = i + 1 | X_n = i\} = 1 - p_i$. No other transitions are allowed in this simplistic model. The transition probability matrix for this DTMC is

$$P = \begin{bmatrix} p_1 & q_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & p_2 & q_2 & 0 & \cdots & 0 & 0 \\ & & \vdots & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & p_{b-1} & q_{b-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

where $q_i = 1 - p_i$. The multi-step shock model is similar to the one-step in that there is zero probability that the process attains a lower cumulative damage level and that there is some probability, $p_{i,i}$, that the process retains its current cumulative damage level. However, in the multi-step model, there exists some probability that a shock will increase the cumulative damage one or more levels. Let $p_{i,j} \equiv P\{X_{n+1} = j | X_n = i\}$, $i \leq j$. Additionally, $\sum_j p_{i,j} = 1$, $0 < p_{i,j} < 1$, $i = 1, \ldots, b-1$, and $j = 1, \ldots, b$.

The transition probabilities for this DTMC are

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,b-1} & p_{1,b} \\ 0 & p_{2,2} & p_{2,3} & \cdots & p_{2,b-1} & p_{2,b} \\ & & \vdots & \ddots & & \\ 0 & 0 & 0 & \cdots & p_{b-1,b-1} & q_{b-1,b} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

With transition probability matrices defined for one-step and multi-step shock models, standard DTMC analyses may be employed to obtain the cumulative distribution function of the random time to first reach the failure state, b. Define α_j as the probability that state j is initially occupied, $\alpha_j = P\{X_0 = j\}$, $j = 1, 2, \ldots, b-1$. We note that $P\{X_0 = b\} = 0$ since the process cannot start in a failed state. The initial probability distribution is defined as $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{b-1} & 0 \end{bmatrix}$. Further define

$$\alpha_j^{(n)} = P\{X_n = j\}, \quad j = 1, 2, \ldots, b$$

where $\alpha_j^{(n)} \geq 0$ and $\sum_{j=1}^{b} \alpha_j^{(n)} = 1$. Letting $\boldsymbol{\alpha}^{(n)} = \begin{bmatrix} \alpha_1^{(n)} & \alpha_2^{(n)} & \cdots & \alpha_b^{(n)} \end{bmatrix}$ and using results from DTMCs, the marginal distribution vector of X_n is

$$\boldsymbol{\alpha}^{(n)} = \boldsymbol{\alpha} P^n, \quad n = 0, 1, 2, \ldots \tag{2.34}$$

Let T_b be a random variable denoting the first time that the DTMC reaches the cumulative damage level b where

$$T_b = \min\{n > 0 : X_n = b\}. \tag{2.35}$$

Let $F(k) \equiv P\{T_b \leq k\}$, $k = 0, 1, 2, \ldots$. The model structure makes it clear that $P\{T_b \leq 0\} = 0$. Suppose $\{T_b < k\}$. Since T_b is the first time the DTMC reaches

the cumulative damage level b, then T_b occurred before time k implying $\{X_k = b\}$. Next, consider the event $\{T_b = k\}$. In this case, the first time to reach the cumulative damage level b occurred at time k, which, by the definition of T_b, is $\{X_k = b\}$. Thus, it is seen that the event $\{T_b \leq k\}$ is equivalent to the event $\{X_k = b\}$ and by Equation (2.34),

$$P\{T_b \leq k\} = \alpha_b^{(k)}.$$

Kolesar [48] provided an example of the multi-step basic shock model utility. While not concerned with finding the lifetime distribution, the author used the shock model formulation to find the conditions and optimal rules for equipment replacement. Valdez-Flores and Feldman [89] also examined a multi-step shock model in optimally selecting partial repair policies for Markov models. In an example problem, the authors provided a slight variation in the multi-step shock model where $p_{i,i} = 0$ for $i = 1, \ldots, b-1$ and $p_{b,b} = 1$. Kasumu and Lešanovský [41] formulated a very similar model to [48] but used a one-step shock model instead to determine the optimal replacement policy.

Extensions of the simplistic version of the shock model with respect to cumulative damage have been formulated. These extensions include the analysis of shock arrival time, shock magnitude, cumulative shocks, and others. Lee and Lee [54] provided another use and extension of the basic one-step shock model by analyzing low-cycle fatigue life prediction under multiaxial loading for 316L stainless steel. In rare cases, it is possible for transitions to occur from state i to state $i-1$ or $i+1$ as

shown in their transition probability matrix

$$P = \begin{bmatrix} p_1 & q_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ r_2 & p_2 & q_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & r_3 & p_3 & q_3 & \cdots & 0 & 0 & 0 \\ & & \vdots & & \ddots & & & \\ 0 & 0 & 0 & & \cdots & r_{b-1} & p_{b-1} & q_{b-1} \\ 0 & 0 & 0 & & \cdots & 0 & 0 & 1 \end{bmatrix}$$

where $r_i+p_i+q_i = 1$. The authors used this shock model to compute and compare the lifetime distributions formed from different angle loadings and provided the loading angle where the material failed quickest. Esary, *et al.* [24] are responsible for one of the earliest and most important papers dealing with shock models and their extension to wear processes. The authors examined the lifetime distribution of a device as a function of the probabilities, $\bar{P}_j$, of surviving j shocks, $j = 0, 1, 2, \ldots$. The device is subject to shocks governed by a Poisson process with lifetime distribution

$$H(t) = 1 - \sum_{j=0}^{\infty} \frac{\bar{P}_j e^{-\lambda t}(\lambda t)^j}{j!}, \qquad t \geq 0 \tag{2.36}$$

where $1 \geq \bar{P}_0 \geq \bar{P}_1 \geq \ldots$ and the probability of failure on the j^{th} shock is $p_j = \bar{P}_{j-1} - \bar{P}_j$. Assuming $1 = \bar{P}_0$, it was shown that $H(t)$ has an increasing failure rate (IFR) if $\bar{P}_j/\bar{P}_{j-1}$ is decreasing in $j = 1, 2, \ldots$. Gottlieb [27] extended the results of [24] by relaxing the assumption of a Poisson damage process. However, he assumed that as the cumulative damage increases, the probability that additional damage will cause failure also increases. Gottlieb [27] provided conditions on the damage process that proves the device's lifetime distribution has an IFR. Shanthikumar and Sumita [80] analyzed a system whose failure was caused by the occurrence of a shock greater than some prespecified level. Associated with their shock model was

a correlated pair (X_n, Y_n) of renewal sequences with joint distribution function

$$F_{X,Y}(x,y) = P\{X_n \leq x, Y_n \leq y\}, \quad n = 0, 1, 2, \ldots. \tag{2.37}$$

Transform results, an exponential limit theorem, and the properties of the failure times were obtained for two models based on the magnitude of the n^{th} shock defined by the random variable X_n. The first model considered X_n's correlation with the length, Y_n, of the interval since the last shock and the second model considered X_n's correlation with the length, Y_n, of the subsequent interval until the next shock. If X_n and Y_n are independent and Y_n are identically and exponentially distributed, the shock model reduces to that in [24]. In [80], the authors assume that X_n and Y_n are correlated. Defining $\{N(t) : t \geq 0\}$ as the counting process associated with Y_n, Shanthikumar and Sumita [80] further define $M(t)$ as the maximum shock that occurs in time t. Given a shock greater than a prespecified level z, the lifetime distribution, T_z, is

$$P\{T_z \leq t\} = P\{M(t) > z\}. \tag{2.38}$$

Let Ω be the sample space for the correlated pair of renewal sequences, (X_n, Y_n). Similar to Equation (2.35), the failure time $T_z(\omega)$, is

$$T_z(\omega) = \inf\{t : M(\omega, t) \geq z\} \tag{2.39}$$

for each sample path $\omega \in \Omega$. For the correlation between shock magnitude and the length of the interval since the last shock, the authors defined $V(z,t) = P\{M(t) \leq z\}$ and $W_z(t) = P\{T_z \leq t\}$. Conditioning on the first renewal time, Y_1, using the regenerative property for (X_n, Y_n), taking the Laplace transform of $V(z,t)$ and using the dual relationship in Equation (2.38), the lifetime distribution, $P\{T_z \leq t\}$, was determined, as well as the first and second moments. However, the marginal dis-

tributions and marginal densities for Equation (2.37) are required to determine the lifetime distribution. The system failure time for the case of correlated shock magnitude and the time interval until the $(n+1)^{th}$ shock was derived in a similar manner. Shanthikumar and Sumita [80] also proved that by assuming $0 < F_{X,Y}(x,y) < 1$, $0 < x < \infty$, $0 < y < \infty$ and $\mathbb{E}[Y] < \infty$,

$$\frac{T_z}{\mathbb{E}[T_z]} \xrightarrow{dist} U \tag{2.40}$$

as $z \to \infty$ where $P\{U \leq x\} = 1 - e^{-x}$. Sumita and Shanthikumar [85] provided similar results to [80] for the lifetime distribution of a system subject to shocks. In [85], however, a system failure occurred when the cumulative magnitude of shocks exceeded a prespecified value z instead of a single shock exceeding some prespecified level in [80]. Furthermore, the lifetime distribution contained in [85] was a 2-dimensional Laplace transform whereas the distribution in [80] was 1-dimensional. Igaki, *et al.* [33] examined a trivariate stochastic process with random shocks, X_n, at random intervals, Y_n, with random system state J_n. Their purpose was to extend the results in [80] and [85] and to incorporate the influences of an environmental process on the correlation between X_n and Y_n as shown by the similarity to Equation (2.37) of the joint distribution function

$$F_{i,j}(x,y) = P\{X_{n+1} \leq x, Y_{n+1} \leq y, J_{n+1} = j | J_n = i\} \tag{2.41}$$

where $\mathbf{F}(x,y) = [F_{i,j}(x,y)]$. The system lifetime distribution and its moments were explicitly derived for two shock models where failure in one is defined by a shock greater than some prespecified level and failure in the other is defined by the cumulative damage reaching a prespecified level. Igaki, *et al.* [33] also proved that if $F_{X,Y}(x,y)$ is replaced by $\mathbf{F}(x,y)$ in the assumptions for Equation (2.40) and $\mathbf{F}(\infty,\infty)$ is an ergodic stochastic matrix, then Equation (2.40) holds for the trivariate process. Waldmann [92] also examined a system subject to shocks and accumulating damage.

He related the shock process to an environmental process described by varying external and internal factors. This work provided sufficient conditions for an optimal replacement decision rule with respect to total cost and some bounds associated with this rule. The probability that the system survives to time n is given by $g(a_n, i_n, x_n)$ and $\{Y_n : n \geq 0\} = 0$ where a_n is the cumulative damage following the decision to replace or not replace the system, i_n is the state of the environment, x_n is the magnitude of the n^{th} shock, and $\{Y_n : n \geq 0\}$ is either 0 or 1, indicating the system is functioning or failed. If the accumulated damage exceeded an environmental state dependent, prespecified critical value, then the system should be replaced. Puri and Singh [74] provided results to find an optimal replacement time based on cost considerations of shocks, maintenance, and replacement. The shock models addressed thus far apply only to instantaneous jumps in the cumulative damage at discrete times. Another methodology that addresses cumulative damage is a wear model.

Wear models differ from shock models in that damage accrues continually over time, without jumps and a failure occurs only when the level of cumulative damage reaches some prespecified value, x. Continuous deterioration and aging are other terms that are normally associated with wear models. The cumulative damage at time t is defined by the stochastic process $\{X(t) : t \geq 0\}$ with no initial damage, i.e. $X(0) = 0$ and is an increasing function. Esary, *et al.* [24] proved that if $\{X(t) : t \geq 0\}$ is a Markov process and if $P\{X(t + \Delta t) - X(t) \leq u | X(t) = z\}$ is decreasing in both z and t for $t \geq 0$, $z \geq 0$, and $\Delta t \geq 0$, then the associated first passage time to reach the cumulative damage value, x,

$$T_x = \inf\{t : (X(t) > x\} \tag{2.42}$$

has an increasing hazard rate average (IHRA) distribution. Oluyede [68] defined F, which is a right-continuous distribution where $F(0+) = 0$, to be an IHRA distribu-

tion if and only if for all $0 \leq \alpha \leq 1$ and $x \geq 0$,

$$\overline{F}(\alpha x) \geq \overline{F}^{\alpha}(x) \tag{2.43}$$

where $\overline{F} = 1 - F$. Additionally, given a nonnegative, non-decreasing function g, F is also an IHRA distribution if and only if

$$\int g(x)dF(x) \leq \left\{ \int g^{\alpha}\left(\frac{x}{\alpha}\right) dF(x) \right\}^{1/\alpha} \tag{2.44}$$

for all $0 \leq \alpha \leq 1$.

Abdel-Hameed [2] examined a device subject to wear and modelled the wear, $\{X(t) : t \geq 0\}$, as an increasing Lévy process [39], which is a stochastic process with stationary, independent increments and continuous in probability for every positive ε. That is,

$$\lim_{s \to t} P\{|X(t) - X(s)| > \varepsilon\} = 0, \quad t \geq 0. \tag{2.45}$$

Abdel-Hameed [2] defined a random failure threshold, Y, and designated the failure time of the device as

$$T_Y = \inf\{t : X(t) > Y\}. \tag{2.46}$$

It was shown that if the Lévy measure μ is finite, then the right-tail probability, G, where $G(x) = P\{Y > x\}$, and the survival probability both have increasing failure rates or both have decreasing failure rates given further assumptions on G. Kharoufeh [44] examined the wear of a single-unit system subject to a random environment modelled as a continuous time Markov chain (CTMC). The failure time distribution and moments were completely specified in the transform space.

Lastly, some models incorporate both shock and wear. Hordijk and Van Der Duyn Schouten [32] incorporated both shock and wear models into their Markov decision drift process. The drift process is defined by the nine-tuple $(S, A_1, A_2, q, \Pi, p, c_1, c_2, f)$ where Π defines the jump distribution which governs the shock model and f is the drift function defining the wear process. The main result of this paper was to provide an average optimal policy. Çinlar [18] also examined the combination of shock and wear models in addition to Markov additive processes and Lévy processes.

For the most part, these analytical models provide little in the area of numerical implementation. If examples are provided, they normally assume specific parameter values. Thus, there is no specified manner to incorporate degradation data into these analytical models. Thus, we turn our attention to examining the methods to obtain the remaining lifetime distribution in reliability theory.

2.4 Failure and Degradation-Based Reliability

Failure-based reliability is used to estimate the lifetime distribution and its parameters when sufficient, complete (and/or censored) failure time data exists. If prior knowledge of the lifetime distribution exists for similar components, then often the lifetime distribution is assumed to follow the same distribution of a similar component. For example, Elsayed [23] wrote about numerous distributions, including the Weibull, exponential, lognormal, and normal, and discussed various components whose failure times were characterized by these distributions. He mentioned the exponential model could best describe components operating at normal conditions and with failure due to a secondary cause while the lognormal could best describe the life data of a semiconductor failure mechanism.

However, if the distribution and/or its parameters are unknown, methods are available to determine the "best" distribution and its parameters. Software packages such as BestFit [8] and ExpertFit [25] provide test statistics on many types of

distribution and estimates of the distribution parameters by fitting the lifetime data to many different, but well-known distributions. BestFit fits up to 27 different distributions while ExpertFit claims 40 distributions. These distributions are ranked in order by the Chi-squared, Kolmogorov-Smirnov, and/or the Anderson-Darling tests. A downfall of these software packages is that only a limited number of distributions can be compared and the failure data may not originate from any of these well-known distribution. Numerous other techniques which account for censored failure data exist but are not examined in this dissertation.

Degradation-based reliability focuses on using measures of component degradation, not failure data, to assess the remaining lifetime of a component. Degradation is also known as cumulative damage. Chao [16] provided an excellent review of degradation topics that included four sets of degradation data, the methodology used to determine shelf lives, the study of growth curves, sigmoids, degradation data collection, and methods to model the degradation process. While little detail is given, he does quickly mention the wear model through Equation (2.42). Additionally, he mentioned that if

$$X(t) = a + bt + W(t) \tag{2.47}$$

where $W(t)$ is Brownian motion, is used to estimate the cumulative damage, then the lifetime distribution is inverse Gaussian.

Meeker and Escobar [59] devoted a chapter to degradation analysis. Their general degradation path model, given the observed sample degradation $y_{i,j}$ of unit i at time t_j, is

$$y_{i,j} = \mathcal{D}_{i,j} + \epsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i \tag{2.48}$$

where $\mathcal{D}_{i,j}$ is the actual path of unit i at time $t_{i,j}$ and $\epsilon_{i,j} \sim N(0, \sigma_\epsilon)$ is residual deviation. Additionally, $\mathcal{D}_{i,j}$ is a function of $t_{i,j}$ and the unknown parameters $\beta_{1,i}, \ldots, \beta_{k,i}$.

While Meeker and Escobar [59] indicated that $\beta_{1,i}, \ldots, \beta_{k,i}$ can be estimated and then employed to predict future degradation of the individual unit, the remainder of their discussion focused on inferences and predictions about the population which required the mean values and covariances of the unknown parameters, $\beta_1, \ldots, \beta_k$. No further insight for estimating the single-unit degradation path was included.

Lu and Meeker [57] reviewed nonlinear regression models and formed a two-stage method to estimate the model parameters. Stage 1 parameter estimates were obtained from each degradation path. These estimates were then transformed, if necessary, to ensure the parameter estimates came from a multivariate normal distribution. All Stage 1 estimates were then combined to determine estimates of the mean, variance, and covariance which were then utilized to find the lifetime distribution. Chan and Meeker [15] incorporated time series modelling to estimate the degradation probability distribution at a given point in time and the lifetime probability distribution for solar reflector material. The degradation was modelled with an autoregressive (AR) process using predicted daily degradation based on data recorded from previous years. Monte Carlo simulation provided numerous sample paths which were used to form empirical distribution functions for the degradation and lifetime distributions.

If lifetime observations and degradation data are not available, accelerated life testing can quickly provide such data. Meeker and Escobar [59] state that highly reliable components motivate the need for accelerated degradation tests (ADT) and accelerated life tests (ALT) which can provide degradation and lifetime data. These tests increase the use-rate, aging-rate, or stress level of a component in an effort to decrease the amount of time required for component degradation and failure. The authors provided numerous strategies and techniques which allow degradation and lifetime information in accelerated tests to be appropriately used in real time analysis. Additionally, Meeker, *et al.* [60] provided a general framework for accelerated tests to better predict life performance in a highly-variable environment.

These accelerated tests, however, require knowledge of the relation between the accelerated variables and accelerated time in order to convey the results to non-accelerated components. Elsayed [23] indicated that model identification and parameter estimation are two important problems with accelerated life tests because there is no method to determine if the model and parameters determined from accelerated life tests are accurate without having actual failure information. Other difficulties [59] with accelerated tests include:

1. Choosing appropriate variables that affect failure,
2. Extraneous failures resulting from high levels of accelerated variables,
3. Extrapolating if more than one failure mode is present, and
4. Verifying acceleration's representation of actual process.

Given these difficulties, accelerated tests may provide an inaccurate representation of the actual process and may not be suitable for certain reliability assessments.

This chapter has examined current applications of systems prognosis, prognosis from state estimation techniques, analytical models providing a closed-form lifetime distribution, and failure-based and degradation-based reliability. No single method or implementation in this chapter provides the capability, without making simplifying assumptions, to incorporate real-time degradation data and to determine the remaining lifetime distribution which enables systems prognosis. In Chapter III, we derive analytical models, examine numerical techniques, and demonstrate through numerical implementation our procedures to obtain the remaining lifetime distribution which does enable systems prognosis.

III. Mathematical Model Description

This chapter provides the general mathematical model description and main results for the cumulative distribution function (CDF) of the time to reach a prespecified degradation threshold using real-time degradation and the associated random environment of a single-unit system. Analytical models for the lifetime distribution provide a viable means for assessing the remaining lifetime of the system within the context of systems prognosis. Hence, our purpose for this chapter is to analyze models that can be numerically implemented for the development of a prognostic capability.

In order to account for the impact of the unit's random environment, the environment is modelled as a finite-state, continuous-time stochastic process $\mathcal{Z} \equiv \{Z(w) : w \geq 0\}$. In particular, we assume that the degradation rate of the single-unit system depends on the state of its random environment and that the time spent in that state has an associated probability distribution. Assuming the system operates to failure without any repairs, the cumulative degradation up to time w may be described by a monotonically increasing stochastic process $\mathcal{X} \equiv \{X(w) : w \geq 0\}$.

Initially, the system is assumed to be in perfect working order but degrades in its random environment until the cumulative degradation of the system exceeds a fixed threshold value x, at which time it fails. The lifetime is denoted by the random variable $T(x)$. The lifetime distribution is our main focus since the remaining lifetime distribution can be determined from the lifetime distribution as will be shown in Section 3.4.

The remainder of this chapter is organized as follows. The next section provides a summary of the assumptions regarding the mathematical model for the Markovian environment. Section 3.2 provides the derivations of the full lifetime distribution when the system is subject to a nonhomogeneous and homogeneous Markovian environment. In Section 3.3, we demonstrate that the lifetime distribution is a matrix-

exponential distribution with easily derived moments. In Section 3.4, we show that the remaining lifetime distribution is obtained directly from the lifetime distribution– a necessary result for enabling systems prognosis. Section 3.5 examines numerical methods to implement the analytical models. It begins by separately examining the estimation techniques associated with observable environments and observable degradation, the data sources utilized to demonstrate these estimation techniques, and the goodness-of-fit test required to test our procedures. This chapter concludes by examining four illustrative examples based upon the two cases of an observable environment and observable unit degradation.

3.1 Markovian Environment

We first consider the case in which the random environment is modelled as a continuous-time Markov chain (CTMC), $\mathcal{Z} \equiv \{Z(w) : w \geq 0\}$. The process has sample space $S = \{1, ..., K\}$, where K is a positive integer, and the random variable $Z(w)$ represents the state of the random environment at time w. Before reviewing the rudimentary concepts of a CTMC, we first review time-homogeneous discrete-time Markov chains (DTMC) as a stochastic process, $\{Y_n : n \geq 0\}$, embedded at transition epochs, as shown in Figure 3.1, with one-step transition probabilities given by

$$\begin{aligned} p_{i,j} &= P\{Y_{n+1} = j | Y_n = i, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0\} \\ &= P\{Y_{n+1} = j | Y_n = i\} \quad \text{(Markov property)} \\ &= P\{Y_1 = j | Y_0 = i\} \qquad \text{(Time homogeneity).} \end{aligned} \tag{3.1}$$

For a time-homogeneous CTMC, Define $q_{i,j}$ as the transition rate from state i to j, $i \neq j$, and

$$q_i = -q_{i,i} = \sum_j q_{i,j} \tag{3.2}$$

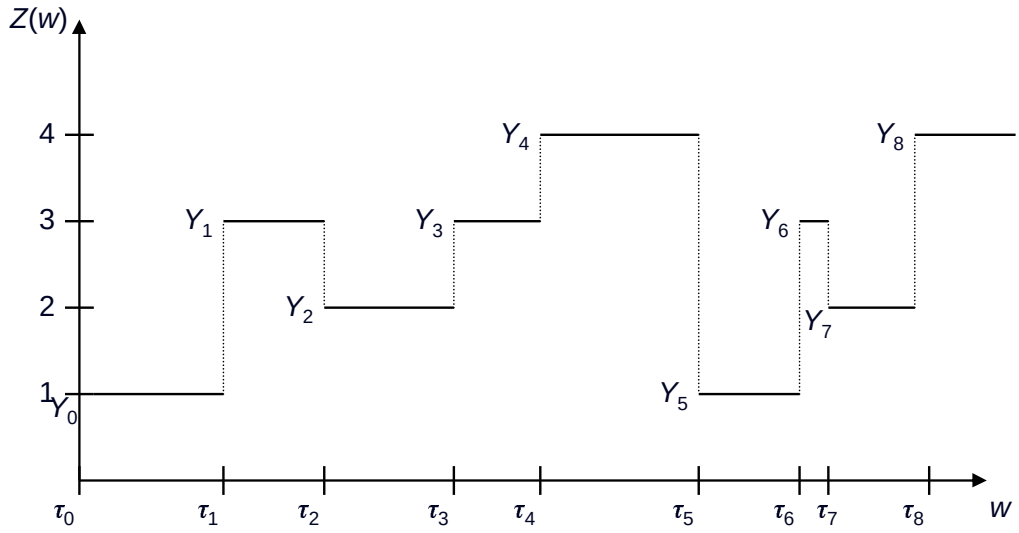

Figure 3.1: Sample path of a continuous-time Markov chain.

where once the system enters state i, it will remain in i for a random amount of time that is distributed exponentially with parameter q_i. The elements, $q_{i,j}$ form the time-homogeneous infinitesimal generator matrix, $\mathbf{Q}$. Define τ_n as the nth transition epoch and let $Y_n \equiv Z(\tau_n^+), n \geq 0$ be the state of the environment just after the nth transition. A sequence of bivariate random variables, $\{(Y_n, \tau_n) : n \geq 0\}$ is a Markov renewal sequence if for all $n \geq 0$,

$$\begin{aligned} \Psi_{i,j}(w) &= P\{\tau_{n+1} - \tau_n \leq w, Y_{n+1} = j | \tau_n, Y_n = i, \tau_{n-1}, Y_{n-1} = i, \ldots, \tau_0, Y_0\} \\ &= P\{\tau_{n+1} - \tau_n \leq w, Y_1 = j | Y_0 = i\} \end{aligned} \tag{3.3}$$

where $\tau_0 = 0, \tau_{n+1} \geq \tau_n$, and $Y_n \in S$. Each of these variables play a role in the definition of a CTMC.

Definition 1. *([50]) If the stochastic process, $\{Z(w) : w \geq 0\}$, has piecewise constant, right continuous sample paths and $\{(Y_n, \tau_n) : n \geq 0\}$ is a Markov renewal sequence that satisfies*

$$\Psi_{i,j}(w) = p_{i,j}\,(1 - e^{q_i w})$$

then $\mathcal{Z}$ is a continuous-time Markov chain.

For a nonhomogeneous continuous-time Markov chain, the holding time distribution may vary over time, but is always exponentially distributed as shown through an example of a three-state infinitesimal generator,

$$\mathbf{Q}(w) = \begin{bmatrix} -2w & w & w \\ 4w & -7w & 3w \\ 2w & w & -3w \end{bmatrix} \tag{3.4}$$

where, for example, the amount of time spent in the second state is exponentially distributed with rate parameter $7w$. With a clear understanding of a homogeneous and nonhomogeneous CTMC and its associated infinitesimal generator matrix, we continue with the development of the more general case of the nonhomogeneous environment.

Uniquely associated with state i of the nonhomogeneous Markov environment is a degradation rate, $r(i)$. The collection of all degradation rates is given by $\mathcal{D} = \{r(1), \ldots, r(K)\}$. If $Z(w) = i \in S$, then the degradation rate, $r(Z(w)) = r(i)$. Additionally, the environment jumps from state $i \in S$ to state $j \in S$ at time w according to the transition probability matrix $\mathbf{P}(v, w) \equiv [p_{i,j}(v, w)]$, where $p_{i,j}(v, w) \equiv P\{Z(w) = j | Z(v) = i\}$. Define $X(w)$ as the cumulative degradation of the system up to time w and let $\mathcal{X} \equiv \{X(w) : w \geq 0\}$ denote the monotonically increasing degradation process defined on $[0, \infty)$. Çinlar [18] indicated that a Markov additive process, $(\mathcal{X}, \mathcal{Z})$, could include a cumulative deterioration process and a Markovian process to find the lifetime distribution under a random threshold. Analysis of the bivariate process, $\{(X(w), Z(w)) : w \geq 0\}$, using a specified failure threshold, is the means by which we obtain the lifetime distribution.

The cumulative degradation up to time $w \geq 0$ is modelled as a Markov reward model of the form

$$X(w) = \int_0^w r(Z(u))du \tag{3.5}$$

where $X(0) \equiv 0$. The system's lifetime expires when the cumulative degradation exceeds a fixed threshold value x. The lifetime distribution of the system is

$$F(x, w) \equiv P\{T(x) \leq w\}. \tag{3.6}$$

Due to the dual relationship

$$T(x) = \inf\{w : X(w) > x\}, \tag{3.7}$$

it follows that

$$F(x, w) = 1 - P\{X(w) \leq x\}. \tag{3.8}$$

Define the joint distribution,

$$V_{i,j}(x, v, w) = P\{X(w) \leq x, Z(w) = j | Z(v) = i\} \tag{3.9}$$

where $V_{i,j}(x, v, w)$ is the joint probability that, at time w, the degradation of the system has not exceeded the failure point x and the environment process is in state $j \in S$ at time w given that the environment was in state $i \in S$ at time v, $v < w$. The matrix $\mathbf{V}(x, v, w) \equiv [V_{i,j}(x, v, w)]$. The row vector $\boldsymbol{\alpha}(v) = [\alpha_1(v) \quad \cdots \quad \alpha_K(v)]$, $\alpha_i(v) \equiv P\{Z(v) = i\}$, is the probability distribution of $\mathcal{Z}$ at time v where

$$\sum_{i=1}^{K} \alpha_i(v) = 1, \tag{3.10}$$

and $\boldsymbol{e}$ is a K-dimensional column vector of ones. Thus, the lifetime distribution may be written as

$$F(x, v, w) = 1 - \boldsymbol{\alpha}(v)\mathbf{V}(x, v, w)\boldsymbol{e}. \tag{3.11}$$

While Equation (3.11) provides the lifetime distribution, it is obvious that we must solve for $V_{i,j}(x, v, w)$ in Equation (3.9) and we begin by showing that $V_{i,j}(x, v, w)$ satisfies a partial differential equation (PDE) in x and w when $\mathcal{Z}$ is a nonhomogeneous, continuous-time Markov chain (CTMC).

3.2 Full Lifetime Distribution

3.2.1 Time-Nonhomogeneous Markov Environment

This subsection considers the case in which the system is subject to a nonhomogeneous Markov environment. We derive a PDE for the $K \times K$ matrix $\mathbf{V}(x, v, w)$ which is then solved via Laplace transforms. The inversion of the transform provides the lifetime distribution as a key step in finding the remaining lifetime distribution. Theorem 1 provides our first main result.

Theorem 1. *If the environment process, $\mathcal{Z}$, with finite-state space S, is a nonhomogeneous, Markov chain with infinitesimal generator matrix, $\mathbf{Q}(w) \equiv [q_{i,j}(w)]$, $i, j \in S$, then the matrix $\mathbf{V}(x, v, w)$ satisfies the matrix PDE*

$$\frac{\partial \mathbf{V}(x, v, w)}{\partial w} + \frac{\partial \mathbf{V}(x, v, w)}{\partial x}\mathbf{R}_D = \mathbf{V}(x, v, w)\mathbf{Q}(w), \tag{3.12}$$

where $\mathbf{R}_D$ is a diagonal matrix of the degradation rates, diag$(r(1), \ldots, r(K))$.

Proof. We prove the theorm by considering both the right- and left-sided limits with respect to the time and spatial variables. For the right-sided limit with respect to the time variable and the left-sided limit with respect to the spatial variable, let $\varepsilon > 0$ denote a small time increment and let $v < w$, where $v, w \in [0, \infty)$.

By our previous definition,

$$V_{i,j}(x, v, w + \varepsilon) \quad = \quad P\{X(w + \varepsilon) \leq x, Z(w + \varepsilon) = j | Z(v) = i\}.$$

Using the law of total probability,

$$\begin{aligned} V_{i,j}(x, v, w + \varepsilon) &= \sum_k P\{X(w + \varepsilon) \leq x, Z(w + \varepsilon) = j | Z(w) = k, Z(v) = i\} \\ &\qquad \times P\{Z(w) = k\} \\ &= \sum_k P\{Z(w + \varepsilon) = j | X(w + \varepsilon) \leq x, Z(w) = k, Z(v) = i\} \\ &\qquad \times P\{X(w + \varepsilon) \leq x | Z(w) = k, Z(v) = i\} P\{Z(w) = k\}. \end{aligned}$$

Since the state of the environment is independent of the cumulative degradation,

$$\begin{aligned} V_{i,j}(x, v, w + \varepsilon) &= \sum_k P\{Z(w + \varepsilon) = j | Z(w) = k, Z(v) = i\} \\ &\qquad \times P\{X(w + \varepsilon) \leq x | Z(w) = k, Z(v) = i\} P\{Z(w) = k\}. \end{aligned} \quad (3.13)$$

By definition, the probability of transitioning from state k to state j at time w is

$$p_{k,j}(w, w + \varepsilon) \quad = \quad P\{Z(w + \varepsilon) = j | Z(w) = k\}, \quad (3.14)$$

and from [47], it is well known that

$$p_{k,j}(w, w + \varepsilon) = \begin{cases} 1 + q_{j,j}(w)\varepsilon + o(\varepsilon), & k = j \\ q_{k,j}(w)\varepsilon + o(\varepsilon), & k \neq j \end{cases} \quad (3.15)$$

where $o(\varepsilon)$ denotes a function that is 'little oh of ε'. That is, a function f is $o(\varepsilon)$ if

$$\lim_{\varepsilon \to 0} \frac{f(\varepsilon)}{\varepsilon} = 0. \quad (3.16)$$

From the Markov property of the $\mathcal{Z}$ process, we have that

$$P\{Z(w+\varepsilon)=j|Z(w)=k,Z(v)=i\}=P\{Z(w+\varepsilon)=j|Z(w)=k\}. \quad (3.17)$$

Applying the Markov property and substituting Equation (3.14), Equation (3.13) becomes

$$\begin{aligned} V_{i,j}(x,v,w+\varepsilon) &= \sum_k p_{k,j}(w,w+\varepsilon)P\{X(w+\varepsilon)\le x|Z(w)=k,Z(v)=i\} \\ &\quad \times P\{Z(w)=k\}. \end{aligned}$$

By unconditioning, we obtain

$$\begin{aligned} V_{i,j}(x,v,w+\varepsilon) &= \sum_k p_{k,j}(w,w+\varepsilon)P\{X(w+\varepsilon)\le x,Z(w)=k|Z(v)=i\} \\ &= \sum_k p_{k,j}(w,w+\varepsilon)V_{i,k}(x-r(k)\varepsilon,v,w). \end{aligned} \quad (3.18)$$

Substituting the case $k=j$, applying Equation (3.15), and simplifying gives

$$\begin{aligned} V_{i,j}(x,v,w+\varepsilon) &= (1+q_{j,j}(w)\varepsilon+o(\varepsilon))\times V_{i,j}(x-r(j)\varepsilon,v,w) \\ &\quad +\sum_{k\neq j} p_{k,j}(w,w+\varepsilon)V_{i,k}(x-r(k)\varepsilon,v,w) \\ &= V_{i,j}(x-r(j)\varepsilon,v,w) \\ &\quad +\varepsilon\sum_k q_{k,j}(w)V_{i,k}(x-r(k)\varepsilon,v,w)+o(\varepsilon). \end{aligned} \quad (3.19)$$

By rearranging the terms of Equation (3.19),

$$V_{i,j}(x,v,w+\varepsilon)-V_{i,j}(x-r(j)\varepsilon,v,w)=\varepsilon\sum_k q_{k,j}(w)V_{i,k}(x-r(k)\varepsilon,v,w)+o(\varepsilon). \quad (3.20)$$

Adding $V_{i,j}(x,v,w) - V_{i,j}(x,v,w)$ to the left side of Equation (3.20) results in

$$[V_{i,j}(x,v,w+\varepsilon) - V_{i,j}(x,v,w)] - [V_{i,j}(x - r(j)\varepsilon, v, w) - V_{i,j}(x,v,w)] \\ = \varepsilon \sum_k q_{k,j}(w) V_{i,k}(x - r(k)\varepsilon, v, w) + o(\varepsilon). \quad (3.21)$$

Dividing by ε and multiplying the second term on the left side of Equation (3.21) by $-r(j)/-r(j)$ gives

$$\frac{V_{i,j}(x,v,w+\varepsilon) - V_{i,j}(x,v,w)}{\varepsilon} - (-r(j)) \frac{V_{i,j}(x - r(j)\varepsilon, v, w) - V_{i,j}(x,v,w)}{-r(j)\varepsilon}$$

while the right side of Equation (3.21) becomes

$$\sum_k q_{k,j}(w) V_{i,k}(x - r(k)\varepsilon, v, w) + \frac{o(\varepsilon)}{\varepsilon}.$$

Now let $\varepsilon \to 0^+$ to obtain the right-sided limit with respect to the time variable and the left-sided limit with respect to the spatial variable shown by

$$\frac{\partial V^+_{i,j}(x,v,w)}{\partial w} + r(j) \frac{\partial V^-_{i,j}(x,v,w)}{\partial x} = \sum_k q_{k,j}(w) V_{i,k}(x,v,w) \quad (3.22)$$

where the superscripts $(+)$ and $(-)$ respectively denote the partial derivative from above and from below (that is, right-sided and left-sided derivatives).

For the left-sided limit of the time variable and the right-sided limit of the spatial variable, let $\varepsilon > 0$ denote a small time increment and let $v < w - \varepsilon < w$, where $v, w \in [0, \infty)$. Assume the environment process is in state j at time w and let $\delta = r(j)\varepsilon$ denote an arbitrarily small degradation increment. Then,

$$V_{i,j}(x + \delta, v, w) \quad = \quad P\{X(w) \leq x + \delta, Z(w) = j | Z(v) = i\}.$$

Using the law of total probability,

$$\begin{aligned} V_{i,j}(x+\delta, v, w) &= \sum_k P\{X(w) \leq x+\delta, Z(w) = j | Z(w-\varepsilon) = k, Z(v) = i\} \\ &\qquad \times P\{Z(w-\varepsilon) = k\} \\ &= \sum_k P\{Z(w) = j | X(w) \leq x+\delta, Z(w-\varepsilon) = k, Z(v) = i\} \\ &\qquad \times P\{X(w) \leq x+\delta | Z(w-\varepsilon) = k, Z(v) = i\} P\{Z(w-\varepsilon) = k\}. \end{aligned}$$

Since the state of the environment is independent of the cumulative degradation,

$$\begin{aligned} V_{i,j}(x+\delta, v, w) &= \sum_k P\{Z(w) = j | Z(w-\varepsilon) = k, Z(v) = i\} \\ &\qquad \times P\{X(w) \leq x+\delta | Z(w-\varepsilon) = k, Z(v) = i\} P\{Z(w-\varepsilon) = k\}. \end{aligned} \tag{3.23}$$

The probability of transitioning from state k to state j at time $w-\varepsilon$ is given by

$$p_{k,j}(w-\varepsilon, w) \;\; = \;\; P\{Z(w) = j | Z(w-\varepsilon) = k\}, \tag{3.24}$$

and

$$p_{k,j}(w-\varepsilon, w) = \begin{cases} 1 + q_{j,j}(w-\varepsilon)\varepsilon + o(\varepsilon), & k = j \\ q_{k,j}(w-\varepsilon)\varepsilon + o(\varepsilon), & k \neq j. \end{cases} \tag{3.25}$$

From the Markov property of the $\mathcal{Z}$ process, we have that

$$P\{Z(w) = j | Z(w-\varepsilon) = k, Z(v) = i\} = P\{Z(w) = j | Z(w-\varepsilon) = k\}. \tag{3.26}$$

Applying Equation (3.26) and substituting Equation (3.24), Equation (3.23) becomes

$$\begin{aligned} V_{i,j}(x+\delta, v, w) &= \sum_k p_{k,j}(w-\varepsilon, w) P\{X(w) \le x+\delta | Z(w-\varepsilon) = k, Z(v) = i\} \\ &\quad \times P\{Z(w-\varepsilon) = k\}. \end{aligned}$$

By unconditioning, we obtain

$$V_{i,j}(x+\delta, v, w) = \sum_k p_{k,j}(w-\varepsilon, w) P\{X(w) \le x+\delta, Z(w-\varepsilon) = k | Z(v) = i\}.$$

For the above summation, let $\delta = r(j)\varepsilon$. Then

$$\begin{aligned} V_{i,j}(x+r(j)\varepsilon, v, w) &= \sum_k p_{k,j}(w-\varepsilon, w) P\{X(w) \le x+r(j)\varepsilon, Z(w-\varepsilon) = k | Z(v) = i\} \\ &= \sum_k p_{k,j}(w-\varepsilon, w) P\{X(w-\varepsilon) \le x + r(j)\varepsilon - r(k)\varepsilon, Z(w-\varepsilon) = k | Z(v) = i\} \\ &= \sum_k p_{k,j}(w-\varepsilon, w) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon). \quad (3.27) \end{aligned}$$

Substituting the case $k = j$, applying Equation (3.25), and simplifying gives

$$\begin{aligned} V_{i,j}(x+r(j)\varepsilon, v, w) &= (1 + q_{j,j}(w-\varepsilon)\varepsilon + o(\varepsilon)) \times V_{i,j}(x, v, w-\varepsilon) \\ &\quad + \sum_{k \ne j} p_{k,j}(w-\varepsilon, w) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon) \\ &= V_{i,j}(x, v, w-\varepsilon) \\ &\quad + \varepsilon \sum_k q_{k,j}(w-\varepsilon) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon) + o(\varepsilon). \quad (3.28) \end{aligned}$$

By rearranging the terms of Equation (3.28),

$$\begin{aligned} &V_{i,j}(x+r(j)\varepsilon, v, w) - V_{i,j}(x, v, w-\varepsilon) \\ &\quad = \varepsilon \sum_k q_{k,j}(w-\varepsilon) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon) + o(\varepsilon). \quad (3.29) \end{aligned}$$

Adding $V_{i,j}(x,v,w) - V_{i,j}(x,v,w)$ to the left side of Equation (3.29) results in

$$\begin{aligned}[V_{i,j}(x + r(j)\varepsilon, v, w) - V_{i,j}(x,v,w)] - [V_{i,j}(x,v,w-\varepsilon) - V_{i,j}(x,v,w)] \\ = \varepsilon \sum_k q_{k,j}(w-\varepsilon) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon) + o(\varepsilon). \quad (3.30)\end{aligned}$$

Dividing by ε, multiplying the first term on the left side of Equation (3.30) by $r(j)/r(j)$, and rewriting the second term on the left side gives the left side as

$$r(j)\frac{V_{i,j}(x + r(j)\varepsilon, v, w) - V_{i,j}(x,v,w)}{r(j)\varepsilon} + \frac{V_{i,j}(x,v,w-\varepsilon) - V_{i,j}(x,v,w)}{-\varepsilon}$$

while the right side of Equation (3.30) becomes

$$\sum_k q_{k,j}(w-\varepsilon) V_{i,k}(x + r(j)\varepsilon - r(k)\varepsilon, v, w-\varepsilon) + \frac{o(\varepsilon)}{\varepsilon}.$$

Now let $\varepsilon \to 0^+$ to obtain the left-sided derivative with respect to the time variable and the right-sided derivative with respect to the spatial variable shown by

$$\frac{\partial V^-_{i,j}(x,v,w)}{\partial w} + r(j)\frac{\partial V^+_{i,j}(x,v,w)}{\partial x} = \sum_k q_{k,j}(w) V_{i,k}(x,v,w). \quad (3.31)$$

By combining Equation (3.22) and (3.31), it is seen that the partial derivatives of $V_{i,j}(x,v,w)$ exist for each i,j and $V_{i,j}(x,v,w)$ satisfies the partial differential equation

$$\frac{\partial V_{i,j}(x,v,w)}{\partial w} + r(j)\frac{\partial V_{i,j}(x,v,w)}{\partial x} = \sum_k q_{k,j}(w) V_{i,k}(x,v,w) \quad (3.32)$$

which may be written in matrix form as

$$\frac{\partial \mathbf{V}(x,v,w)}{\partial w} + \frac{\partial \mathbf{V}(x,v,w)}{\partial x}\mathbf{R}_D = \mathbf{V}(x,v,w)\mathbf{Q}(w), \quad (3.33)$$

□

Equation (3.12) forms a system of PDEs that must be solved simultaneously. For example, if the environment has two states (i.e. $K = 2$), then there are four PDEs to solve with four unknown functions. Suppressing the arguments of $V_{i,j}, i, j \in S$, these PDEs are

$$\begin{aligned}
\frac{\partial V_{1,1}}{\partial w} + r(1)\frac{\partial V_{1,1}}{\partial x} &= q_{1,1}(w)V_{1,1} + q_{2,1}(w)V_{1,2} \\
\frac{\partial V_{1,2}}{\partial w} + r(2)\frac{\partial V_{1,2}}{\partial x} &= q_{1,2}(w)V_{1,1} + q_{2,2}(w)V_{1,2} \\
\frac{\partial V_{2,1}}{\partial w} + r(1)\frac{\partial V_{2,1}}{\partial x} &= q_{1,1}(w)V_{2,1} + q_{2,1}(w)V_{2,2} \\
\frac{\partial V_{2,2}}{\partial w} + r(2)\frac{\partial V_{2,2}}{\partial x} &= q_{1,2}(w)V_{2,1} + q_{2,2}(w)V_{2,2}.
\end{aligned} \tag{3.34}$$

In general, if the environment has K states, there are K^2 PDEs with K^2 unknown functions. Since x and w are weakly coupled, the method of separation of variables cannot be used to solve this system of equations. Thus, we resort to a Laplace transform approach.

By definition (*cf.* [67]), the Laplace transform of an absolutely integrable function, $h(w)$, is

$$h^*(s) = \int_0^\infty e^{-sw} h(w) dw, \qquad Re(s) > 0 \tag{3.35}$$

and the Laplace transform of its derivative, $h'(w)$, which is also absolutely integrable is

$$\mathcal{L}(h'(w)) \equiv \int_0^\infty e^{-sw} h'(w) dw = sh^*(s) - h(0). \tag{3.36}$$

Additionally, it is well-known (*cf.* [50]) that the relationship between the Laplace transform and Laplace-Stieltjes transform (LST) is

$$\widetilde{h}(s) = sh^*(s) \tag{3.37}$$

where the LST of h, denoted by $\widetilde{h}(s)$, is given by

$$\widetilde{h}(s) = \int_0^\infty e^{-sw} dh(w), \quad Re(s) > 0. \tag{3.38}$$

We now show that solving the PDE in Equation (3.12) provides the LST of the lifetime distribution given by

$$\widetilde{F}(u, v, w) = \int_0^\infty e^{-ux} dF(x, v, w). \tag{3.39}$$

Theorem 2. *Suppose the system is subject to a nonhomogeneous, finite Markov environment process $\mathcal{Z}$ whose distribution at time v is $\boldsymbol{\alpha}(v)$, has infinitesimal generator matrix, $\mathbf{Q}(w)$, and degradation rate matrix, $\mathbf{R}_D$. The Laplace-Stieltjes transform of the lifetime distribution, with respect to x, is*

$$\widetilde{F}(u, v, w) = 1 - \boldsymbol{\alpha}(v) \exp\left(\int_v^w (\mathbf{Q}(\tau) - u\mathbf{R}_D) d\tau \right) \boldsymbol{e}, \quad Re(u) > 0. \tag{3.40}$$

Proof. Assume that $\mathbf{V}(x, v, w)$ is absolutely integrable in x for each v and w and that $\partial \mathbf{V}(x, v, w)/\partial w$ is also absolutely integrable. Define

$$\mathbf{V}^*(u, v, w) = \int_0^\infty e^{-ux} \mathbf{V}(x, v, w) dx \tag{3.41}$$

as the matrix Laplace transform of $\mathbf{V}(x, v, w)$ with respect to x. Taking the Laplace transform of both sides of Equation (3.12) yields

$$\frac{\partial \mathbf{V}^*(u, v, w)}{\partial w} + (u\mathbf{V}^*(u, v, w) - \mathbf{V}(0, v, w))\mathbf{R}_D = \mathbf{V}^*(u, v, w)\mathbf{Q}(w). \tag{3.42}$$

Applying the initial condition, $\mathbf{V}(0, v, w) = \mathbf{0}$, and simplifying, the ordinary differential equation (ODE) in w yields

$$\frac{d\mathbf{V}^*(u, v, w)}{dw} + \mathbf{V}^*(u, v, w)(u\mathbf{R}_D - \mathbf{Q}(w)) = \mathbf{0}. \tag{3.43}$$

This ODE is not exact and we must find an integrating factor to obtain its solution [49]. Thus, multiplying each term by the integrating factor on the right

$$\boldsymbol{\mu}(w) = \exp\left(\int_0^w (u\mathbf{R}_D - \mathbf{Q}(w))dw\right), \tag{3.44}$$

and simplifying, provides

$$\frac{d}{dw}\left[\mathbf{V}^*(u, v, w)\exp\left(\int_0^w (u\mathbf{R}_D - \mathbf{Q}(w))dw\right)\right] = \mathbf{0}. \tag{3.45}$$

After integrating with respect to w,

$$\mathbf{V}^*(u, v, w)\exp\left(\int_0^w (u\mathbf{R}_D - \mathbf{Q}(w))dw\right) = \mathbf{C} \tag{3.46}$$

where the matrix $\mathbf{C}$ is $u^{-1}\mathbf{I}$ after applying the initial condition, $\mathbf{V}^*(u, v, 0) = u^{-1}\mathbf{I}$. Thus,

$$\mathbf{V}^*(u, v, w)\exp\left(\int_0^w (u\mathbf{R}_D - \mathbf{Q}(w))dw\right) = u^{-1}\mathbf{I}, \tag{3.47}$$

and

$$\begin{aligned}\mathbf{V}^*(u,v,w) &= u^{-1}\left[\exp\left(\int_0^w (u\mathbf{R}_D - \mathbf{Q}(w))dw\right)\right]^{-1} \\ &= u^{-1}\exp\left(\int_0^w (\mathbf{Q}(w) - u\mathbf{R}_D)dw\right). \end{aligned} \tag{3.48}$$

Since this derivation ranges from v to w, we further simplify to

$$\mathbf{V}^*(u,v,w) = u^{-1}\exp\left(\int_v^w (\mathbf{Q}(\tau) - u\mathbf{R}_D)d\tau\right), \tag{3.49}$$

and by taking the LST of Equations (3.11) and (3.49), we obtain,

$$\widetilde{F}(u,v,w) = 1 - \boldsymbol{\alpha}(v)\widetilde{\mathbf{V}}(u,v,w)\boldsymbol{e}.$$

□

While this result provides the analytical lifetime distribution of a system subject to a nonhomogeneous Markov environment, the procedures required to numerically implement this distribution are nontrivial. Specifically, estimating the time variant generator matrix, $\mathbf{Q}(\tau)$, requires an enormous amount of data. Moreover, there is no guarantee that the integral of Equation (3.49) can be evaluated. Under certain conditions of the generator matrix, we demonstrate that we can obtain the moments for the lifetime distribution in Equation (3.40) in Section 3.3.2. Since this analytical result is not viable for systems prognosis, we turn our attention to time homogeneous environments.

3.2.2 Time-Homogeneous Markov Environment

In this subsection, we impose a more restrictive assumption by assuming a homogeneous Markovian environment. We retain the Markov property of the $\mathcal{Z}$

process

$$P\{Z(w+\varepsilon)=j|Z(w)=k,Z(v)=i\}=P\{Z(w+\varepsilon)=j|Z(w)=k\},\ \ v<w$$

which shows the history of the process before time w is not relevant. We also impose time homogeneity which mathematically implies

$$P\{Z(w+\varepsilon)=j|Z(w)=k\}=P\{Z(\varepsilon)=j|Z(0)=k\},\ \text{for all}\ \varepsilon>0,$$

or $p_{k,j}(w+\varepsilon,w)=p_{k,j}(\varepsilon)$, where the distribution of the $\mathcal{Z}$ process over any two non-overlapping intervals of length ε are the same. For the nonhomogeneous Markovian environment, this is not the case and the distributions may differ while preserving the Markov property. Additionally, homogeneity removes the time dependency of the generator matrix such that $\mathbf{Q}(\tau)=\mathbf{Q}$, for all τ, and $\boldsymbol{\alpha}(v)$ becomes $\boldsymbol{\alpha}=\begin{bmatrix}\alpha_1 & \cdots & \alpha_K\end{bmatrix}, \alpha_i\equiv P\{Z(0)=i\}$. Thus, we drop the dependency of the joint distribution $V_{i,j}(x,v,w)$ on v and simply write

$$V_{i,j}(x,w)=P\{X(w)\leq x,Z(w)=j|Z(0)=i\}.$$

Additionally, the LST of the lifetime distribution, $F(x,w)$ assuming a homogeneous Markovian environment will be denoted by

$$\widetilde{F}(u,w)=\int_0^\infty e^{-ux}dF(x,w). \tag{3.50}$$

The next theorem provides an analytical result for the LST of the lifetime distribution.

Theorem 3. *Suppose the system is subject to a homogeneous finite Markovian environment process $\mathcal{Z}$ with initial distribution, $\boldsymbol{\alpha}$, infinitesimal generator $\mathbf{Q}$, and degradation rate matrix, $\mathbf{R}_D$. The Laplace-Stieltjes transform of the lifetime distribution*

is

$$\widetilde{F}(u,w) = 1 - \boldsymbol{\alpha}\exp((\mathbf{Q} - u\mathbf{R}_D)w)\boldsymbol{e}, \quad Re(u) > 0. \tag{3.51}$$

Proof. Let $\mathbf{Q}(\tau) = \mathbf{Q}$ and $v = 0$. Then Equation (3.40) becomes

$$\begin{aligned}
\widetilde{F}(u,w) &= 1 - \boldsymbol{\alpha}\exp\left(\int_0^w (\mathbf{Q} - u\mathbf{R}_D)d\tau\right)\boldsymbol{e} \\
&= 1 - \boldsymbol{\alpha}\exp\left((\mathbf{Q} - u\mathbf{R}_D)\,\tau\Big|_{\tau=0}^{\tau=w}\right)\boldsymbol{e} \\
&= 1 - \boldsymbol{\alpha}\exp((\mathbf{Q} - u\mathbf{R}_D)w)\boldsymbol{e}
\end{aligned}$$

□

The integral in this proof is legitimate since the integral of a matrix is a matrix of integrals of the individual elements in the matrix.

The same result, expressed as the Laplace transform of the lifetime distribution, $F(x,w)$, is derived by Kulkarni, *et al.* [51]. However, the authors used a purely probabilistic approach and obtained the distribution as a two-dimensional transform. They inverted the LST with respect to w to obtain the Laplace transform of $F(x,w)$ with respect to x. We began with a nonhomogeneous, joint probability distribution and found a PDE that was then changed into an ODE. This ODE was solved via transforms to obtain the one-dimensional Laplace-Stieltjes transform of the lifetime distribution, $F(x,v,w)$. Imposing homogeneity, we obtained the Laplace-Stieltjes transform of $F(x,w)$.

Equations (3.40) and (3.51) have a particular structure that allows for further analysis. Specifically, we show in the next section that these two equations, under certain conditions for the nonhomogeneous environment, are matrix-exponential distributions that lead to simple analytical moment expressions.

3.3 Moments of the Lifetime Distribution

The transforms of the lifetime distributions, Equations (3.40) and (3.51), are expressed in such a way that obtaining the moments are straightforward in most cases. Let Y denote a positive random variable with cumulative distribution function (CDF) $F(\cdot)$ whose LST is $\widetilde{F}(s)$ and let the nth moment of Y be denoted $\mathbb{E}[Y^n]$. The transform method for computing the nth moment of a distribution is through the nth order differentiation of the LST such that

$$\mathbb{E}[Y^n] = (-1)^n \frac{d^n \widetilde{F}(s)}{ds^n}\bigg|_{s=0} \tag{3.52}$$

where $d^n\widetilde{F}(s)/ds^n$ is the nth derivative of the $\widetilde{F}(s)$ with respect to s. However, this equation cannot be utilized to find the moments of the distributions in Equations (3.40) and (3.51) because the matrices in the matrix exponential do not commute. In other words, $\mathbf{R}_D\mathbf{Q} \neq \mathbf{Q}\mathbf{R}_D$ in general. Thus, we must resort to other methods in order to obtain the moments. This section explores the matrix-exponential distribution which allows us to find the LST of the moments in a simpler fashion than that shown in [44] and [46].

3.3.1 Matrix-Exponential (ME) Distributions

Asmussen and Bladt [6] define a matrix-exponential (ME) distribution as a probability distribution on $[0, \infty)$ with a rational Laplace transform. Let Y_A denote a positive random variable with CDF $A(w) = P\{Y_A \leq w\}$. The distribution $A(\cdot)$ is a ME distribution with representation ME$(\boldsymbol{\alpha}, \mathbf{T}, \boldsymbol{s})$ if

$$A(w) = 1 + \boldsymbol{\alpha} \exp(\mathbf{T}w)\mathbf{T}^{-1}\boldsymbol{s} \tag{3.53}$$

where $\boldsymbol{\alpha}$ is a row vector, $\mathbf{T}$ is a real or complex valued matrix, and $\boldsymbol{s}$ is a column vector. While the authors note that it may be difficult to determine when a triple,

$(\boldsymbol{\alpha}, \mathbf{T}, \boldsymbol{s})$, designates a ME distribution, it is necessary that $\boldsymbol{\alpha}\mathbf{T}^{-1}\boldsymbol{s} = 1$ and $\boldsymbol{\alpha}\boldsymbol{s} \geq 0$. Given a ME distribution, the moments are easily determined. Asmussen and Bladt [6] give the nth moment as

$$\mathbb{E}[Y_A^n] = (-1)^{n+1} n! \boldsymbol{\alpha}\mathbf{T}^{-n-1}\boldsymbol{s}. \tag{3.54}$$

Bladt and Neuts [9] rewrite Equation (3.53) by letting $\boldsymbol{s} = -\mathbf{T}\boldsymbol{e}$ because indicate it is always possible to choose $\boldsymbol{s} = -\mathbf{T}\boldsymbol{e}$ (see [6]). Making this substitution results in the ME distribution

$$A(w) = 1 - \boldsymbol{\alpha} \exp(\mathbf{T}w)\boldsymbol{e} \tag{3.55}$$

with representation ME$(\boldsymbol{\alpha}, \mathbf{T})$ and nth moment

$$\mathbb{E}[Y_A^n] = (-1)^n n! \boldsymbol{\alpha}\mathbf{T}^{-n}\boldsymbol{e}. \tag{3.56}$$

3.3.2 ME Distribution Moments

Suppose a single-unit system is subject to a time-nonhomogeneous Markovian environment process, $\mathcal{Z}$, with lifetime distribution as shown in Equation (3.40). The following definitions and theorems, summarized from Olmsted [69], define an analytic function used in our next result.

Definition 2. *A power series in $(\tau - a)$ is a series of the form*

$$\sum_{n=0}^{\infty} a_n (\tau - a)^n \tag{3.57}$$

where $\{a_n\}$ and a are constants.

Theorem 4. *If a power series in* $(\tau - a)$ *has a positive radius of convergence, and if*

$$f(\tau) \equiv \sum_{n=0}^{\infty} a_n (\tau - a)^n \tag{3.58}$$

is in the interval of convergence, then $f(\tau)$ *is continuous and has continuous derivatives throughout that interval and*

$$a_n = \frac{f^{(n)}(a)}{n!}, \quad n \geq 0 \tag{3.59}$$

where $f^{(n)}(a)$ *is the nth-order derivative of* $f(\tau)$ *at* a*. The Taylor series for* $f(\tau)$ *is*

$$f(\tau) = f(a) + f'(a)(\tau - a) + \frac{f''(a)}{2!}(\tau - a)^2 + \dots. \tag{3.60}$$

Definition 3. *A function* f *is analytic at* $\tau = a$ *if and only if it has a Taylor series at* $\tau = a$ *which represents the function in some neighborhood of* $\tau = a$*.*

Theorem 5. *The sum of analytic functions is analytic if division by zero is not involved.*

The following result shows that the moments of a component's lifetime may be computed if its nonhomogeneous infinitesimal generator matrix consists of analytic functions.

Theorem 6. *If* $q_{i,j}(\tau)$ *is an analytic function for* $i, j \in S, i \neq j$*, then*

$$\widetilde{F}(u, v, w) = 1 - \boldsymbol{\alpha} \exp\left(\mathbf{T}(u)\,(w - v)\right) \boldsymbol{e}, \quad Re(u) > 0,$$

is a ME($\boldsymbol{\alpha}, \mathbf{T}(u)$*) distribution whose nth moment* $(n \geq 1)$ *is*

$$\mathbb{E}[T^n(u)] = (-1)^n n! \boldsymbol{\alpha} \mathbf{T}(u)^{-n} \boldsymbol{e}, \quad Re(u) > 0, \tag{3.61}$$

where

$$\mathbf{T}(u) = \begin{bmatrix} O_{1,1}(v,w) - ur(1) & \cdots & O_{1,K}(v,w) \\ O_{2,1}(v,w) & \cdots & O_{2,K}(v,w) \\ \vdots & \ddots & \vdots \\ O_{K,1}(v,w) & \cdots & O_{K,K}(v,w) - ur(K) \end{bmatrix}.$$

and $O_{i,j}(v,w)$ is the remaining series after $(w-v)$ is factored from the power series integration.

Proof. Let $q_{i,j}(\tau)$ be an analytic function. Then,

$$\int_v^w (\mathbf{Q}(\tau) - u\mathbf{R}_D)d\tau = \begin{bmatrix} \int_v^w (q_{1,1}(\tau) - ur(1))d\tau & \cdots & \int_v^w q_{1,K}(\tau)d\tau \\ \int_v^w q_{2,1}(\tau)d\tau & \cdots & \int_v^w q_{2,K}(\tau)d\tau \\ \vdots & \ddots & \vdots \\ \int_v^w q_{K,1}(\tau)d\tau & \cdots & \int_v^w (q_{K,K}(\tau) - ur(K))d\tau \end{bmatrix}$$

where each $q_{i,j}(\tau)$ has a power series representation including $q_{i,i}(\tau)$ because

$$q_{i,i}(\tau) = -\sum_j q_{i,j}(\tau).$$

Hence, each $\int_v^w q_{i,j}(\tau)d\tau$ results in a collection of $(w^p - v^p)$ terms, p is a positive integer, with associated coefficients. The diagonals have an additional term of $-(w-v)ur(i)$. Whether p is odd or even, $(w-v)$ can be factored out of each of the $(w^p - v^p)$ terms. Let $O_{i,j}(v,w)$ be the remaining series once $(w-v)$ is factored. Then,

$$\int_v^w (\mathbf{Q}(\tau) - u\mathbf{R}_D)d\tau = \begin{bmatrix} O_{1,1}(v,w) - ur(1) & \cdots & O_{1,K}(v,w) \\ O_{2,1}(v,w) & \cdots & O_{2,K}(v,w) \\ \vdots & \ddots & \vdots \\ O_{K,1}(v,w) & \cdots & O_{K,K}(v,w) - ur(K) \end{bmatrix} (w-v).$$

Let

$$\mathbf{T}(u) = \begin{bmatrix} O_{1,1}(v,w) - ur(1) & \cdots & O_{1,K}(v,w) \\ O_{2,1}(v,w) & \cdots & O_{2,K}(v,w) \\ \vdots & \ddots & \vdots \\ O_{K,1}(v,w) & \cdots & O_{K,K}(v,w) - ur(K) \end{bmatrix}.$$

Then

$$\widetilde{F}(u,v,w) = 1 - \boldsymbol{\alpha} \exp\left(\mathbf{T}(u)(w-v)\right) \boldsymbol{e}. \tag{3.62}$$

□

Without Equation (3.56), we would be required to use Equation (3.52) for which the lifetime moments in the nonhomogeneous Markovian environment are not easily computed. As an example, let

$$\mathbf{Q}(\tau) = \begin{bmatrix} -3\tau & 3\tau \\ 2\tau & -2\tau \end{bmatrix}, \tag{3.63}$$

$$\mathbf{R}_D = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}. \tag{3.64}$$

It is clear that $\mathbf{Q}(\tau)$ is analytic. Then

$$\begin{aligned}
\int_v^w (\mathbf{Q}(\tau) - u\mathbf{R}_D)d\tau &= \int_v^w \begin{bmatrix} -3\tau - u & 3\tau \\ 2\tau & -2\tau - 4u \end{bmatrix} d\tau \\
&= \begin{bmatrix} -\frac{3}{2}\tau^2 - u\tau & \frac{3}{2}\tau^2 \\ \tau^2 & -\tau^2 - 4u\tau \end{bmatrix}\Bigg|_v^w \\
&= \begin{bmatrix} -\frac{3}{2}w^2 - uw & \frac{3}{2}w^2 \\ w^2 & -w^2 - 4uw \end{bmatrix} - \begin{bmatrix} -\frac{3}{2}v^2 - uv & \frac{3}{2}v^2 \\ v^2 & -v^2 - 4uv \end{bmatrix} \\
&= \begin{bmatrix} -\frac{3}{2}w^2 - uw + \frac{3}{2}v^2 + uv & \frac{3}{2}w^2 - \frac{3}{2}v^2 \\ w^2 - v^2 & -w^2 - 4uw + v^2 + 4uv \end{bmatrix} \\
&= \begin{bmatrix} -(\frac{3}{2}w + \frac{3}{2}v + u) & \frac{3}{2}(w + v) \\ w + v & -(w + v + 4u) \end{bmatrix} (w - v).
\end{aligned}$$

Thus,

$$\widetilde{F}(u, v, w) = 1 - \boldsymbol{\alpha} \exp\left(\begin{bmatrix} -(\frac{3}{2}w + \frac{3}{2}v + u) & \frac{3}{2}(w + v) \\ w + v & -(w + v + 4u) \end{bmatrix} (w - v) \right) \boldsymbol{e},\ Re(u) > 0,$$

and the nth moment computed through the numerical inversion (with respect to u) of the Laplace-Stieltjes transform

$$\mathbb{E}[T^n(u)] = n!\boldsymbol{\alpha} \begin{bmatrix} -(\frac{3}{2}w + \frac{3}{2}v + u) & \frac{3}{2}(w + v) \\ w + v & -(w + v + 4u) \end{bmatrix}^{-n} \boldsymbol{e},\quad Re(u) > 0.$$

It is easily shown that the one-dimensional lifetime distribution for the homogeneous Markovian environment shown in Equation (3.51) is a ME distribution since it satisfies $\boldsymbol{\alpha}\mathbf{T}^{-1}\boldsymbol{s} = 1$ and $\boldsymbol{\alpha}\boldsymbol{s} \geq 0$.

Proposition 1. *Suppose the single-unit system is subject to a homogeneous Markov environment process $\mathcal{Z}$ with initial distribution, $\boldsymbol{\alpha}$, infinitesimal generator $\mathbf{Q}$, and*

degradation rate matrix, $\mathbf{R}_D$*. The Laplace-Stieltjes transform of the lifetime distribution,*

$$\widetilde{F}(u,w) = 1 - \boldsymbol{\alpha}\exp((\mathbf{Q} - u\mathbf{R}_D)w)\boldsymbol{e}, \quad Re(u) > 0,$$

is a ME$(\boldsymbol{\alpha}, \mathbf{Q} - u\mathbf{R}_D)$ *distribution with moments*

$$\mathbb{E}[T^n(u)] = (-1)^n n!\boldsymbol{\alpha}(\mathbf{Q} - u\mathbf{R}_D)^{-n}\boldsymbol{e}, \quad Re(u) > 0. \tag{3.65}$$

This result highlights the ease with which the moments for any ME distribution can be computed. The transform of Equation (3.65) can be inverted numerically using the one-dimensional inversion algorithm of Abate and Whitt [1]. Equation (3.65) can also be compared with the results of Kharoufeh [44] and Kharoufeh and Sipe [46] who derived moments for their two-dimensional and one-dimensional lifetime distributions using Equation (3.52). First, to show these results are equivalent, note that

$$\begin{aligned}
\mathbb{E}[T^n(u)] &= (-1)^n n!\boldsymbol{\alpha}(\mathbf{Q} - u\mathbf{R}_D)^{-n}\boldsymbol{e} \\
&= (-1)^n n!\boldsymbol{\alpha}((-1)(u\mathbf{R}_D - \mathbf{Q}))^{-n}\boldsymbol{e} \\
&= (-1)^n n!\boldsymbol{\alpha}(-1)^{-n}(u\mathbf{R}_D - \mathbf{Q})^{-n}\boldsymbol{e} \\
&= n!\boldsymbol{\alpha}(u\mathbf{R}_D - \mathbf{Q})^{-n}\boldsymbol{e}
\end{aligned} \tag{3.66}$$

which is the same result provided in [44] and [46]. Thus, concluding that the lifetime distribution is a ME distribution provides a significant advantage over previous methods for determining the moments of the lifetime distribution. With the lifetime distribution and its moments derived, we now shift our attention to deriving the remaining lifetime distribution which ultimately leads to the ability to implement and perform system prognosis.

3.4 Remaining Lifetime Distribution

Thus far we have derived analytical models of the lifetime distribution of a system subject to nonhomogeneous and homogeneous Markov environments and have shown that we can easily find the moments, specifically for the homogeneous case. We have also noted that it would be very difficult to numerically implement the lifetime distribution assuming a nonhomogeneous environment. Thus, we turn our attention solely to the homogeneous environment. In this section we show that if we have the full lifetime distribution then we also have the remaining lifetime distribution which is critical for performing systems prognosis.

Assume that the system has not failed by time ξ_0. The remaining lifetime distribution is given by

$$\begin{aligned} R(x, w|\xi_0) &= P\{T(x) \geq w + \xi_0 | T(x) > \xi_0\} \\ &= \frac{P\{T(x) \geq w + \xi_0, T(x) > \xi_0\}}{P\{T(x) > \xi_0\}} \\ &= \frac{P\{T(x) \geq w + \xi_0\}}{P\{T(x) > \xi_0\}} \\ &= \frac{1 - F(x, w + \xi_0)}{1 - F(x, \xi_0)} \\ &= \frac{\boldsymbol{\alpha}\mathbf{V}(x, w + \xi_0)\boldsymbol{e}}{\boldsymbol{\alpha}\mathbf{V}(x, \xi_0)\boldsymbol{e}}. \end{aligned} \tag{3.67}$$

From Equation (3.67), it is obvious that the remaining lifetime distribution is obtainable if the full lifetime distribution is known. By the method of integrating the tail, the nth moment of the lifetime distribution, for $n \geq 1$, is given by

$$\begin{aligned} m^{(n)}(x) \equiv \mathbb{E}[T^n(x)] &= \int_0^\infty n w^{n-1} P\{T(x) > w\} dw \\ &= \int_0^\infty n w^{n-1} \boldsymbol{\alpha}\mathbf{V}(x, w)\boldsymbol{e} dw, \end{aligned} \tag{3.68}$$

and the nth moment of the unit's remaining lifetime, given that it has survived at least ξ_0 time units, is

$$\begin{aligned} m^{(n)}(x|\xi_0) &\equiv \mathbb{E}[T^n(x)|T(x) > \xi_0] \\ &= \int_0^\infty nw^{n-1}R(x,w|\xi_0)dw \\ &= \frac{1}{\boldsymbol{\alpha}\mathbf{V}(x,\xi_0)\boldsymbol{e}}\int_{\xi_0}^\infty nw^{n-1}\boldsymbol{\alpha}\mathbf{V}(x,w)\boldsymbol{e}dw. \end{aligned} \tag{3.69}$$

While we derived closed-form expressions for the Laplace-Stieltjes transform of the moments of the lifetime, these are not readily available for the remaining lifetime distributions. However, the mean remaining lifetime may be obtained if

$$m^{(1)}(x|\xi_0) = \frac{1}{\boldsymbol{\alpha}\mathbf{V}(x,\xi_0)\boldsymbol{e}}\int_{\xi_0}^\infty \boldsymbol{\alpha}\mathbf{V}(x,w)\boldsymbol{e}dw \tag{3.70}$$

is evaluated numerically.

While the analytical models derived in this chapter do provide the full and remaining lifetime distributions, the incorporation of real-time degradation data into these analytical models must be developed in order to provide a capability to perform systems prognosis. We show in the remainder of this dissertation that it is possible to numerically incorporate degradation data from a system subject to homogeneous environments. We first examine numerical methods associated with each parameter in the lifetime distribution based upon a homogeneous continuous-time Markov chain (CTMC) and demonstrate in Chapter IV that we can further generalize the CTMC by relaxing the Markov assumption.

3.5 Numerical Methods

This section provides the numerical methods, data sources, goodness-of-fit tests and examples required to demonstrate that the remaining lifetime distribution for systems prognosis is viable and numerically implementable. A basic assumption,

however, is that data obtained from optimally-located sensors provides the information required to implement the techniques in this section. There are four numerical examples which increase in complexity in order to demonstrate our procedures.

For the homogeneous Markovian environment, we examine two scenarios to find the lifetime distribution given in Equation (3.51): an observable environment and observable degradation. The observable environment scenario requires estimation of the generator matrix $\mathbf{Q}$, whereas the observable degradation scenario requires estimation of $\mathbf{Q}$, the initial probability distribution vector ($\boldsymbol{\alpha}$), the diagonal matrix of degradation rates ($\mathbf{R}_D$), and the number of environment states, K. Having this information allows us to characterize the probability distribution and moments of the random lifetime as well as the remaining lifetime distribution. We begin by exploring estimation techniques for the observable environment, specifically the infinitesimal generator matrix and then explore estimation techniques for the observable degradation scenario, specifically the initial probability distribution, the diagonal matrix of degradation rates, and the number of environment states, K. We examine the use of two data sources and briefly explore the statistical tests used to determine the viability of the estimation techniques via the accuracy of the lifetime distribution itself.

3.5.1 Observable Environment

We first assume that the environment process $\{Z(w) : w \geq 0\}$ is continuously observable over the time interval $[0, T]$. The system has not failed by time T, and began its lifetime in perfect working order at time 0 (i.e., $X(0) \equiv 0$). The environment is continuously observed up to time T so that, at each transition epoch, we record the current and subsequent states of the random environment. Some additional notation will make the procedure more transparent. Let $q_{i,j}$ denote the rate at which the environment transitions from state $i \in S$ to state $j \in S$, $j \neq i$. Since we assume a homogeneous Markovian environment process $\{Z(w) : w \geq 0\}$, the holding time in

state i, given that the subsequent state is $j(j \neq i)$, is exponentially distributed with rate parameter $q_{i,j}$. Let $N_{i,j}(T)$ denote the random, integer number of transitions of the process from i to j in a time interval of length T. Moreover, let $H_i(T)$ denote the total holding time in state $i \in S$ during $[0, T]$. It is known [3, 7] that

$$q_{i,j} = \lim_{T \to \infty} \frac{N_{i,j}(T)}{H_i(T)}. \tag{3.71}$$

Therefore, for sufficiently large T, we may approximate $q_{i,j}$, $j \neq i$ by

$$q_{i,j} \approx \hat{q}_{i,j}(T) = \frac{N_{i,j}(T)}{H_i(T)}. \tag{3.72}$$

The diagonal elements of the generator matrix are obtained as

$$\hat{q}_{i,i}(T) = -\sum_{j \neq i} \hat{q}_{i,j}(T) \quad i \in S. \tag{3.73}$$

We note that, for a fixed observation interval, it is also possible to estimate the generator matrix by observing k independent sample paths of the $\mathcal{Z}$ process. Define $N_{i,j}^{(k)}$ as the total number of transitions from state i to j over all k trials and define $H_i^{(k)}$ as the total holding time in state i over all k trials. The off-diagonal elements of the generator matrix are given by

$$q_{i,j} = \lim_{k \to \infty} \frac{N_{i,j}^{(k)}}{H_i^{(k)}} \tag{3.74}$$

so that for k independent sample paths, the estimate of $q_{i,j}$, $j \neq i$ is

$$q_{i,j} \approx \hat{q}_{i,j}^{(k)} = \frac{N_{i,j}^{(k)}}{H_i^{(k)}}. \tag{3.75}$$

We obtain the diagonal elements by

$$\hat{q}_{i,i}^{(k)} = -\sum_{j \neq i} \hat{q}_{i,j}^{(k)} \quad i \in S. \tag{3.76}$$

It has been shown in [7] that

$$q_{i,j} = \lim_{T \to \infty} \frac{N_{i,j}(T)}{H_i(T)} = \lim_{k \to \infty} \frac{N_{i,j}^{(k)}}{H_i^{(k)}}. \tag{3.77}$$

In the examples, we show that we can adequately estimate $q_{i,j}$ for the observable environment scenario. Thus, the only requirement for systems prognosis in an observable environment is the estimation of the generator matrix which we demonstrate in Section 3.5.5. While this is a valid scenario, the most difficult scenario is clearly having only degradation data. This next scenario requires additional estimation techniques in the case when only the degradation of the system is observable.

3.5.2 Observable Degradation

In this subsection, we find that all of the estimation techniques hinge upon the determination of the degradation rates. We rely upon the assumption that there is a one-to-one relationship between the number of degradation rates and the number of environment states. Thus, the collection of the initial degradation rates of each degradation path provides the capability to find the initial probability distribution, and some estimate of the number of degradation rates provides an estimate for the number of environment states. Additionally, we can estimate when the degradation rates change which provides the number of transitions and the time spent in the associated environment state so that the generator matrix can be estimated. We now examine each of these techniques in detail and provide a step-by-step procedure at the end of this subsection for the estimation of these parameters in the observable degradation scenario.

To estimate the degradation rates, we approximate the true degradation path using simple, piecewise linear functions that pass through a finite number of observations. The slope of each line segment approximates the degradation rate over a given time interval. Let $\mathcal{T} \equiv \{t_j : j = 0, 1, 2, \ldots, M\}$. For each observation time $t_j \in \mathcal{T}$, we approximate the potential degradation rates of the process via

$$\gamma_j \equiv \frac{\hat{X}(t_j) - \hat{X}(t_{j-1})}{t_j - t_{j-1}}, \quad j = 1, 2, \ldots, M. \tag{3.78}$$

Additionally, we assume that the length of the time interval between two observations $(t_j - t_{j-1})$ is fixed for all intervals and we gather these observed rates in the set $\Gamma \equiv \{\gamma_j : j = 1, 2, \ldots, M\}$ over the interval $[t_0, t_M]$. We then perform K-means cluster analysis to obtain K distinct environment states. The centroid of each cluster corresponds to the mean degradation rate for each state of the environment. While this method may be elementary, we demonstrate its effectiveness in the examples.

The techniques to estimate the number of environment states can be somewhat subjective. If the true lifetime distribution is known, numerical measures that use some criteria to calculate the distance between the true and approximated distributions may be used to find the best value of K over some range of K. Conover [19] provides some useful measures, such as the Cramér-von Mises test two-sample statistic and the Kolmogorov-Smirnov (KS) test statistic, to accomplish this task. However, since the true distribution is unknown and sought, such approaches are not applicable. It may be possible to obtain a good estimate of K by relying on the experience of subject matter experts (SMEs) or through an evaluation of the degradation rates via cluster analysis.

The opinions of subject matter experts are frequently used, but cannot be relied upon in all circumstances. The experts may potentially miss environment states which have not been observed or overestimate the number of states if some can be combined due to little difference in the degradation rates. Thus, we must rely

upon an assumed one-to-one relationship between the number of environment states and the number of degradation rates. From this assumption, the determination of the number of degradation rates also provides the number of environment states.

In light of this discussion, cluster analysis techniques, using the observed degradation data, appear to hold the most promise for determining the number of degradation rates, and thus, the appropriate number of environment states. Cluster analysis techniques consist of hierarchical and nonhierarchical methods that attempt to determine the best partitioning of the data. In what follows, we discuss the reasons for using nonhierarchical techniques rather than hierarchical ones.

Hierarchical cluster analysis consists of agglomerative and divisive methods. The agglomerative method initially places each data point into its own cluster. Then using some distance measure, e.g. Euclidean distance, the 'closest' set of data are combined and the newly formed cluster is represented by its mean. This process continues until only a single cluster remains. Divisive methods, on the other hand, initially consider the entire data set as a single cluster. Then, using some distance measure, e.g. farthest neighbor, the single cluster is split into two clusters. This process continues until the number of clusters equals the number of data points. A dendrogram or tree structure (see Figure 3.2) is the standard graphical technique used to analyze the results of hierarchical cluster analysis.

The dendrogram displays combinations of clusters. However, if this technique is used for more than 50 data points, the graph can become difficult to read and interpret. With large data sets, it may be more appropriate to display only the last 30 clusters as shown in Figure 3.2 which combines 2000 observations.

Hierarchical clustering techniques work well if the number of data observations are small. While we encountered a significant time requirement to process over 5000 observations, hierarchical clustering techniques have other drawbacks. They do not use relocation strategies which allow reassignment of data points to another cluster. Additionally, some criteria based upon the distance between the clusters must be

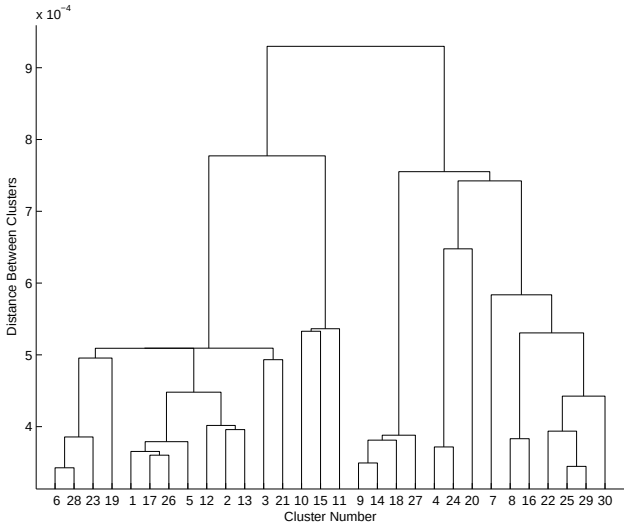

Figure 3.2: An example of a dendrogram using 2000 observations.

developed to find the best partition. This criteria 'cuts' the dendrogram to provide the number of clusters. For example, adding a cut line to Figure 3.2 partitions the dendrogram into three clusters as shown in Figure 3.3. Since the data used

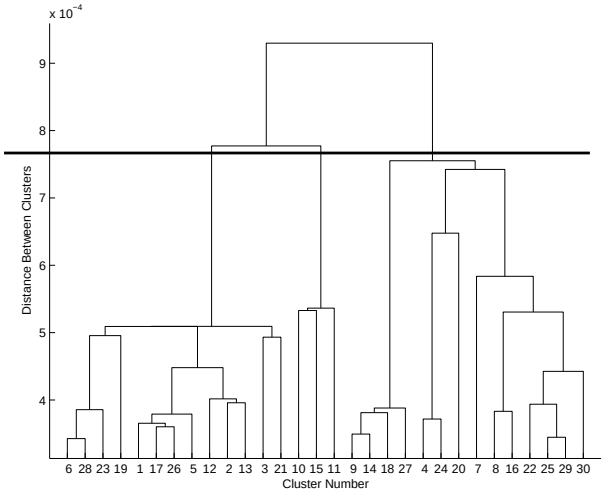

Figure 3.3: Dendrogram with three clusters.

for this dendrogram were generated from three environment states, partitioning the dendrogram into three clusters was appropriate. However, of the 2000 data points, 1882 data points would fall into the first cluster, 31 in the second cluster, and 87 in the

final cluster which does not well represent the actual breakout. Lastly, many of the hierarchical methods are adversely affected by outliers [87]. With these undesirable effects, nonhierarchical clustering techniques, such as K-means cluster analysis, were considered as an alternative method for determining an appropriate estimate of the number of environment states. For our purposes, we assume the number of clusters is equivalent to the number of environment states.

A key component of most cluster analysis techniques is the determination of the number of clusters in a set of observations. Milligan and Cooper [61] used known data sets (with known number of clusters) and compared 30 clustering techniques to determine how each technique performed on these data sets. While no technique found the correct number of clusters for all data sets, Milligan and Cooper [61] indicated that Calinski and Harabasz' [14] method performed better than the others. Hence, we use this technique to estimate the number of environment states. Timm [87] labeled this technique the pseudo F-statistic which has the value F_K and is plotted for $2, 3, \ldots, K$. Additionally, Calinski and Harabasz [14] indicated that the absolute maximum, the first local maximum, or a point with a comparatively rapid increase designates the best choice for the number of clusters. The authors also stated that if the value F_K is monotonically increasing in K, then the number of clusters should equal the number of observations.

Assuming there are K clusters, the technique compares K-means via one-way analysis of variance (ANOVA). Designate $\bar{y}_i$ as the mean of each cluster i, $i = 1, 2, \ldots, K$, and n_i as the number of observations in the ith cluster where

$$N = \sum_{i=1}^{K} n_i. \tag{3.79}$$

and

$$\bar{y} = \frac{1}{K} \sum_{i=1}^{K} \bar{y}_i \tag{3.80}$$

is the overall sample average of degradation rates. Define the sum of squares between clusters, SSB, as

$$SSB = \sum_{i=1}^{K} n_i(\bar{y}_i - \bar{y})^2. \tag{3.81}$$

and the sum of squares within clusters, SSW, as

$$SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_i)^2. \tag{3.82}$$

where $y_{i,j}$ is the j^{th} degradation rate observation in the i^{th} cluster. The F-ratio for K clusters, $2 \leq K < N$, is

$$F_K = \frac{SSB/(K-1)}{SSW/(N-K)}. \tag{3.83}$$

Our objective is to find the "best" value of K via Equation (3.83), over some set $\mathcal{K}$, where F_K is the absolute maximum, the first local maximum, or a point at which the function exhibits a comparatively rapid increase. Let $\mathcal{K} \equiv \{2, 3, \ldots, J\}$ where J is a positive integer (greater than 2). For the sake of computational expedience, it is ideal to choose the smallest value of K that leads to an appropriate representation of the underlying process. As a first resort, we apply the second criterion suggested in [14], choosing that value of K corresponding to the first local maximum over $\mathcal{K}$ given by

$$\hat{K} = \min\{K \in \mathcal{K} : F_K > F_{K+1}\}. \tag{3.84}$$

However, in case F_K is strictly increasing in K, we resort to the remaining two criteria to estimate the smallest possible value. For example, consider a data set with 100,000 observations from ten clusters. The K-means clustering method was applied with $K = 2, 3, \ldots, 15$ and the F_K values were computed for each K. These values are plotted in Figure 3.4 and indicate either eleven clusters using the absolute

maximum F_K value, five clusters if using the first local maximum, and eleven clusters for a point with a comparatively rapid increase. Since one of our concerns is a small sample space, we specifically focus on the first local maximum as our method to estimate the number of environment states and provide examples of this technique later in the dissertation. This technique appears to best fulfill our requirement for the determination of the number of environment states.

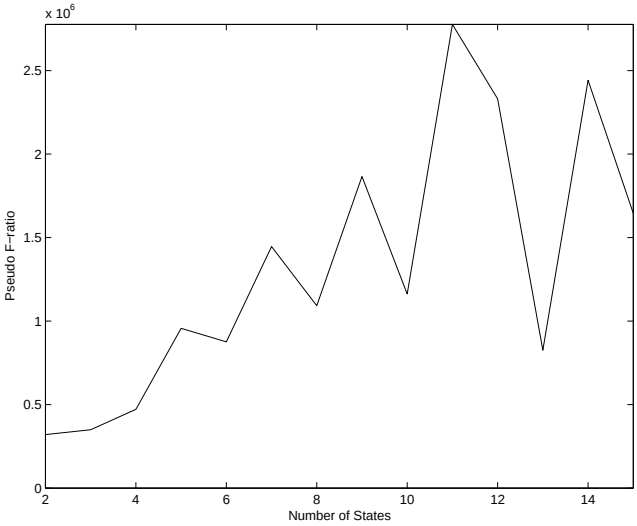

Figure 3.4: Values for F_K showing 10 or 11 clusters.

The following steps provide a synopsis of the estimation techniques for the observable degradation scenario. These estimated parameters can then be utilized to approximate the lifetime distribution and ultimately, the remaining lifetime distribution for use in systems prognosis.

Step 0: *Initialization.*

At time $t_0 \equiv 0$, observe $\hat{X}(t_0) \equiv x_0$.

Step 1: *Observe degradation measures.*

Observe the degradation at times $t_1 < \cdots < t_M$, $M \in \mathbb{N}$ and form the set of observation times

$$\mathcal{T} \equiv \{t_j : j = 0, 1, 2, \ldots, M\}. \tag{3.85}$$

with observations $\hat{X}(t_j)$, $j = 0, 1, 2, \ldots, M$. It is assumed that, at time t_M, the single-unit system has not failed.

Step 2: *Approximate the failure path.*

After observing the degradation path up to time epoch t_M, approximate the true failure path by a simple piecewise-linear approximation that connects the observed degradation measures for each element of $\mathcal{T}$.

Step 3: *Approximate degradation rates via finite difference methods.*

For each observation time $t_j \in \mathcal{T}$, approximate the degradation rates of the process by the difference equation

$$\gamma_j \equiv \frac{\hat{X}(t_j) - \hat{X}(t_{j-1})}{t_j - t_{j-1}}, \quad j = 1, 2, \ldots, M. \tag{3.86}$$

For simplicity, we assume that the length of the time interval between two observations $(t_j - t_{j-1})$ is fixed for all intervals. For the discrete sampling interval, $[t_0, t_M]$, gather the observed rates in a set,

$$\Gamma \equiv \{\gamma_j : j = 1, 2, \ldots, M\}.$$

Step 4: *Compute the $\hat{K}$ distinct degradation rates.*

Select an appropriate value $\hat{K}$ via Equation (3.84) applied to observations in the set Γ. This may be accomplished by using any standard statistical software package. Define $\mathcal{C} = \{C_1, C_2, \ldots, C_{\hat{K}}\}$ as the set of $\hat{K}$ distinct clusters such that $C_i \cap C_j = \emptyset$, $j \neq i$, and μ_i denotes the centroid of cluster C_i, $i = 1, 2, \ldots, \hat{K}$. Each $\gamma_j \in \Gamma$ is therefore assigned to exactly one cluster in $\mathcal{C}$ such that the estimated degradation rate of environment state $j \in S$ is $\hat{r}(j) \equiv \mu_j$.

Step 5: *Construct new degradation path.*

Construct a new piecewise linear degradation path by replacing each $\gamma_j \in \Gamma$

by $\hat{r}(j) \equiv \mu_j$. (Note: if $\hat{r}(j) = \hat{r}(j+1)$ for some j, then the two adjacent line segments are replaced by a single line segment with slope equal to $\hat{r}(j)$.)

Step 6: *Approximate generator matrix.*

Using the piecewise linear estimate of the degradation path of Step 5, estimate the $\hat{K}$-dimensional generator matrix $\mathbf{Q}$ by $\hat{\mathbf{Q}}_{t_M}$ using Equations (3.72) and (3.73).

3.5.3 Data Sources and Experimentation

This section examines both the simulated and empirical data used to evaluate the estimation techniques discussed in this section. For the simulated data, we define numerous infinitesimal generator matrices with between two and fifteen states as well as twenty states. Additionally, we define degradation rate matrices and initial probability distributions when required. Since the amount of time spent in any state is assumed to be exponentially distributed, and the system is degrading at a specific rate while in that state, we generate numerous sample degradation paths so that the six steps of subsection 3.5.1 may be applied. For example, Figure 3.5 shows five degradation sample paths simulated from a known generator and degradation rate matrix.

By simulating these degradation paths, we can compare the lifetime distributions resulting from estimated generator matrices and degradation rate matrices with lifetime distributions resulting from the known generator matrices and degradation matrices. The "closeness" of the distributions is statistically evaluated via goodness-of-fit tests discussed in the next section.

For the empirical data tests, we obtained a data set from Virkler, et al. [90] which contains observations of the number of cycles required to propagate a crack to a given length in 2024-T3 aluminum alloy. Sixty-eight (68) sample paths were used to form an empirical lifetime distribution at various crack lengths (failure thresholds). While this empirical distribution cannot provide the true lifetime distribution, it does

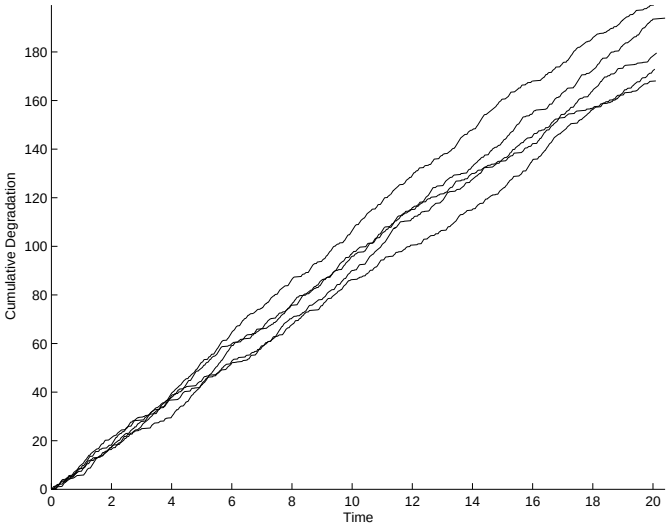

Figure 3.5: A sample of five linear degradation paths.

provide a distribution to test our procedure against using standard goodness-of-fit tests. A graphical depiction of all 68 sample paths is shown in Figure 3.6.

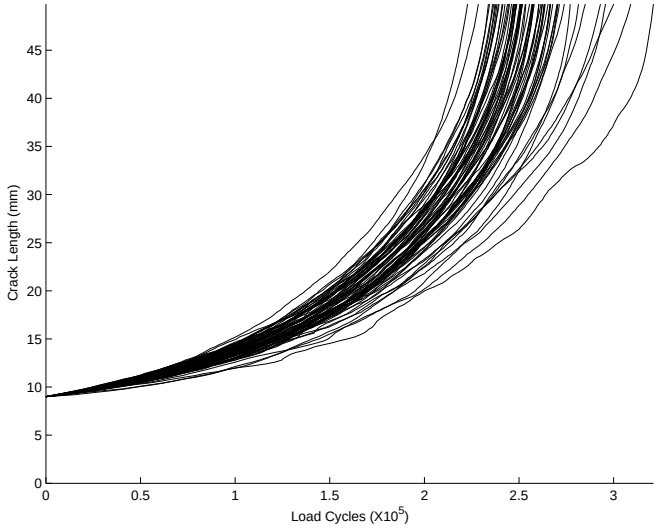

Figure 3.6: Crack propagation for 68 samples of 2024-T3 aluminum alloy.

3.5.4 Goodness-of-Fit Test

This section provides the theoretical foundation for the goodness-of-fit test used to compare two distributions throughout the remaining text. Using known

parameters in one lifetime distribution and estimated parameters from Sections 3.5.1 and 3.5.2 in the other lifetime distribution, this test provides the basis to indicate whether these two distributions are statistically equivalent, and as such, is critical to the veracity of the results in the next subsection.

The Cramér-von Mises test statistic [19] is used to compare two empirical CDFs. The null hypothesis that two distribution functions, U_1 and U_2, are equivalent is

$$H_0 : U_1(x) = U_2(x), \text{ for all } x \in (-\infty, \infty)$$

and the alternative hypothesis is

$$H_1 : U_1(x) \neq U_2(x), \text{ for at least one value of } x.$$

The test statistic, κ_2, with empirical distribution functions U_1 and U_2 evaluated at $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$ is defined as

$$\kappa_2 = \frac{rs}{(r+s)^2} \left\{ \sum_{i=1}^{r} [U_1(x_i) - U_2(x_i)]^2 + \sum_{j=1}^{s} [U_1(y_j) - U_2(y_j)]^2 \right\}. \tag{3.87}$$

For the scenarios we examine, we compare the CDFs, one from the known process and one from the estimated process, at m fixed points, $\tau_1, \tau_2, \ldots, \tau_m$. Since $r = s$ and $x_i = y_i = \tau_i$, the appropriate test statistic simplifies as

$$\kappa_2 = \frac{1}{2} \sum_{i=1}^{m} [U_1(\tau_i) - U_2(\tau_i)]^2 . \tag{3.88}$$

We denote the critical value by κ^*. If $\kappa_2 < \kappa^*$, then we fail to reject the null hypothesis that $U_1 \equiv U_2$. We now have all the required elements to adequately estimate and test the procedure.

3.5.5 Illustrative Examples

This section demonstrates that we can implement the numerical techniques discussed earlier and shows that we can obtain accurate approximations of the lifetime and remaining lifetime distributions. We first show that we can accurately estimate the infinitesimal generator matrix, based upon exponentially distributed holding times, the degradation rates, and the number of environment state. We then show that we can accurately approximate the lifetime distribution which provides the remaining lifetime distribution. We examine four cases, where for each case, we compare, via goodness-of-fit tests, the approximated distribution resulting from estimated parameters and the true distribution resulting from known parameters. Both distributions are formed via Equation (3.51). The four cases are summarized as follows:

Case I. Observable Environment.

Estimate the elements of the generator matrix $\mathbf{Q}$ using simulated data.

Case II. Observable Degradation.

Estimate the elements of the generator matrix $\mathbf{Q}$, the degradation rates, $r(i), i = 1, 2, \ldots, K$, and the initial probability distribution $\boldsymbol{\alpha}$ using simulated data.

Case III. Observable Degradation.

Estimate the elements of the generator matrix, $\mathbf{Q}$, the degradation rates, $r(i), i = 1, 2, \ldots, K$, the initial probability distribution $\boldsymbol{\alpha}$, and the number of environment states using simulated data.

Case IV. Observable Degradation.

Estimate the elements of the generator matrix, $\mathbf{Q}$, the degradation rates, $r(i), i = 1, 2, \ldots, K$, the initial probability distribution $\boldsymbol{\alpha}$, and the number of environment states using real data.

The primary purpose in each case is to compare the lifetime distributions formed by inversion of the Laplace-Stieltjes transform in Equation (3.51). The first

two cases compare the known lifetime distribution resulting from known parameters and the approximated lifetime distribution resulting from known and estimated parameters. The third case compares the known lifetime distribution resulting from known parameters and the approximated lifetime distribution resulting from estimated parameters whereas the fourth case compares an empirical lifetime distribution resulting from 68 observations and the approximated lifetime distribution resulting from estimated parameters. Additional details and some remaining lifetime distribution comparisons are provided in each case.

Case I. Observable Environment

For each $j \in S$, there exists a known degradation rate for the single-unit system, namely $r(j)$. The environment evolves over time and is observed over the interval $[0, T]$. We provide three test cases, namely when the environment process $(\mathcal{Z})$ assumes values on a state space with cardinality 2, 5 and 10, respectively. For each test case, the elements of the generator matrix $\mathbf{Q}$ were drawn from a continuous uniform population on $(20, 40)$. That is, for $j \neq i$,

$$q_{i,j} \sim U(20, 40) \quad i, j \in S.$$

In the usual way, the diagonal elements of the generator matrix are

$$q_{i,i} = -\sum_{j \neq i} q_{i,j} \quad i \in S.$$

The associated degradation rates $(r(j),\ j = 1, 2, \ldots, K)$ were drawn from continuous uniform populations on the interval $(20, 80)$. In all cases, the failure threshold value was fixed at $x = 1$. Moreover, the true generator matrix $\mathbf{Q}$ is estimated by the matrix $\hat{\mathbf{Q}}_T$ with $T = 100, 500, 5000$ using Equations (3.72) and (3.73). Unconditional and residual cumulative probability values were computed using Equation

(3.51) (and Equation (3.67)) using a variant of the numerical inversion algorithm of Moorthy [62]. We compare the CDFs, created by the true and estimated processes, at m fixed points, $\tau_1, \tau_2, \ldots, \tau_m$. For this experiment, and those that follow, goodness-of-fit tests were conducted (at the 0.05 level) to compare the CDFs.

For a 2-state CTMC, the degradation rates, generator matrix and generator matrix estimates at times $T = 100, 500$, and 5000 are

$$\begin{aligned}
\mathcal{D} &= \begin{bmatrix} 34.81 & 78.23 \end{bmatrix} \\
\mathbf{Q} &= \begin{bmatrix} -34.64 & 34.64 \\ 22.01 & -22.01 \end{bmatrix} \\
\hat{\mathbf{Q}}_{100} &= \begin{bmatrix} -30.71 & 30.71 \\ 24.26 & -24.26 \end{bmatrix} \\
\hat{\mathbf{Q}}_{500} &= \begin{bmatrix} -31.71 & 31.71 \\ 19.61 & -19.61 \end{bmatrix} \\
\hat{\mathbf{Q}}_{5000} &= \begin{bmatrix} -34.99 & 34.99 \\ 22.03 & -22.03 \end{bmatrix}
\end{aligned}$$

For the 5-state CTMC, the degradation rates and generator matrices are

$$\mathcal{D} = \begin{bmatrix} 32.26 & 44.88 & 28.66 & 70.33 & 37.48 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} -131.34 & 30.24 & 38.73 & 24.42 & 37.96 \\ 32.64 & -110.57 & 31.87 & 24.00 & 22.07 \\ 29.87 & 30.78 & -128.44 & 30.11 & 37.69 \\ 24.05 & 21.07 & 23.04 & -105.27 & 37.11 \\ 33.00 & 37.66 & 39.53 & 36.46 & -146.66 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{100} = \begin{bmatrix} -119.64 & 22.16 & 29.29 & 25.80 & 42.40 \\ 20.59 & -106.97 & 19.52 & 12.90 & 53.96 \\ 29.41 & 37.99 & -123.14 & 17.20 & 38.54 \\ 32.35 & 18.99 & 42.30 & -124.48 & 30.83 \\ 32.35 & 25.33 & 39.05 & 25.80 & -122.52 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{500} = \begin{bmatrix} -118.69 & 31.21 & 29.02 & 26.64 & 31.82 \\ 25.35 & -126.10 & 38.49 & 19.37 & 42.89 \\ 42.85 & 29.10 & -142.86 & 26.64 & 44.28 \\ 16.90 & 24.34 & 32.57 & -109.78 & 35.98 \\ 35.00 & 21.69 & 38.49 & 36.32 & -131.50 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{5000} = \begin{bmatrix} -121.19 & 32.39 & 29.23 & 24.94 & 34.63 \\ 30.62 & -116.07 & 27.54 & 19.96 & 37.95 \\ 36.15 & 30.81 & -127.07 & 21.91 & 38.20 \\ 26.57 & 25.71 & 30.98 & -119.74 & 36.48 \\ 38.02 & 22.08 & 38.59 & 37.98 & -136.67 \end{bmatrix}$$

For the 10-state CTMC, the degradation rates and generator matrices are

$$\mathcal{D} = \begin{bmatrix} 38.72 & 47.73 & 55.07 & 71.24 & 79.16 & 38.12 & 72.08 & 65.69 & 69.47 & 25.99 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} -249.09 & 20.91 & 20.12 & 27.11 & 30.55 & 24.70 & 33.65 & 20.97 & 36.87 & 34.21 \\ 30.29 & -291.05 & 24.21 & 33.20 & 32.69 & 31.36 & 39.88 & 38.15 & 39.75 & 21.52 \\ 21.51 & 27.76 & -281.16 & 36.86 & 30.70 & 29.77 & 35.24 & 36.04 & 27.07 & 36.22 \\ 30.76 & 25.78 & 24.77 & -249.99 & 25.99 & 20.13 & 33.48 & 27.31 & 28.13 & 33.65 \\ 22.15 & 38.18 & 32.17 & 35.72 & -274.54 & 23.16 & 39.61 & 38.23 & 22.12 & 23.20 \\ 37.14 & 24.15 & 29.79 & 36.24 & 30.43 & -272.03 & 27.12 & 24.24 & 28.09 & 34.84 \\ 24.95 & 29.01 & 31.68 & 38.61 & 26.67 & 25.17 & -263.45 & 27.46 & 33.35 & 26.55 \\ 24.43 & 33.32 & 24.25 & 32.97 & 28.96 & 28.70 & 26.05 & -255.71 & 23.04 & 33.98 \\ 30.68 & 30.71 & 38.00 & 28.50 & 24.74 & 25.24 & 38.05 & 33.92 & -274.66 & 24.80 \\ 26.42 & 36.67 & 23.84 & 26.82 & 31.37 & 34.29 & 28.71 & 28.23 & 38.44 & -274.79 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{100} = \begin{bmatrix} -251.26 & 25.63 & 17.37 & 49.96 & 20.49 & 20.21 & 9.35 & 38.98 & 41.79 & 27.48 \\ 15.93 & -280.64 & 28.95 & 24.98 & 56.33 & 26.95 & 42.06 & 7.80 & 50.15 & 27.48 \\ 21.24 & 25.63 & -290.09 & 14.99 & 46.09 & 20.21 & 28.04 & 31.18 & 75.23 & 27.48 \\ 31.86 & 32.04 & 34.75 & -315.24 & 30.73 & 47.16 & 28.04 & 46.77 & 50.15 & 13.74 \\ 15.93 & 57.67 & 34.75 & 19.99 & -321.07 & 47.16 & 28.04 & 38.98 & 16.72 & 61.84 \\ 26.55 & 25.63 & 23.16 & 24.98 & 10.24 & -227.67 & 18.69 & 38.98 & 25.08 & 34.35 \\ 37.17 & 38.45 & 40.54 & 24.98 & 46.09 & 6.74 & -280.68 & 31.18 & 41.79 & 13.74 \\ 0.00 & 32.04 & 23.16 & 14.99 & 15.36 & 26.95 & 28.04 & -228.02 & 66.87 & 20.61 \\ 37.17 & 57.67 & 17.37 & 54.96 & 10.24 & 20.21 & 23.36 & 7.80 & -276.89 & 48.10 \\ 31.86 & 12.82 & 46.33 & 19.99 & 25.61 & 26.95 & 9.35 & 38.98 & 33.43 & -245.30 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{500} = \begin{bmatrix} -253.89 & 30.98 & 23.87 & 31.51 & 18.57 & 32.27 & 29.55 & 19.84 & 37.87 & 29.42 \\ 23.69 & -257.95 & 32.82 & 32.56 & 27.86 & 28.06 & 27.36 & 28.66 & 22.99 & 33.95 \\ 19.17 & 26.02 & -234.36 & 25.21 & 32.50 & 25.25 & 31.74 & 20.95 & 29.76 & 23.76 \\ 30.45 & 28.50 & 31.33 & -272.35 & 30.18 & 26.66 & 43.78 & 20.95 & 37.87 & 22.63 \\ 28.20 & 24.78 & 34.31 & 27.31 & -278.84 & 28.06 & 26.27 & 40.79 & 35.17 & 33.95 \\ 22.56 & 19.83 & 17.90 & 15.76 & 31.34 & -214.58 & 28.46 & 23.15 & 21.64 & 33.95 \\ 34.97 & 34.70 & 44.76 & 33.61 & 48.75 & 42.09 & -316.51 & 19.84 & 35.17 & 22.63 \\ 12.41 & 42.13 & 43.27 & 21.01 & 42.95 & 16.84 & 27.36 & -261.02 & 39.22 & 15.84 \\ 37.22 & 35.94 & 23.87 & 18.91 & 22.05 & 25.25 & 32.83 & 24.25 & -254.28 & 33.95 \\ 28.20 & 24.78 & 44.76 & 28.36 & 13.93 & 32.27 & 33.93 & 33.07 & 31.11 & -270.41 \end{bmatrix}$$

$$\hat{\mathbf{Q}}_{5000} = \begin{bmatrix} -243.61 & 30.08 & 20.45 & 30.69 & 20.55 & 35.77 & 24.75 & 25.80 & 30.13 & 25.39 \\ 20.62 & -261.76 & 27.75 & 24.39 & 39.39 & 25.40 & 26.53 & 35.88 & 27.33 & 34.49 \\ 22.85 & 22.99 & -245.86 & 24.39 & 34.08 & 29.65 & 32.72 & 22.52 & 33.87 & 22.79 \\ 27.70 & 34.81 & 38.64 & -288.45 & 31.16 & 33.77 & 41.21 & 27.27 & 28.49 & 25.39 \\ 30.93 & 30.61 & 24.03 & 24.80 & -250.59 & 29.25 & 26.21 & 27.50 & 23.36 & 33.89 \\ 24.59 & 31.13 & 29.61 & 19.22 & 24.53 & -243.31 & 26.32 & 27.61 & 26.16 & 34.13 \\ 33.04 & 37.70 & 32.53 & 33.38 & 44.29 & 32.18 & -308.65 & 25.12 & 40.17 & 30.23 \\ 21.24 & 38.36 & 34.92 & 25.11 & 39.92 & 22.07 & 28.00 & -269.60 & 32.58 & 27.40 \\ 36.64 & 38.49 & 28.95 & 30.18 & 19.23 & 25.40 & 32.72 & 24.11 & -274.91 & 39.21 \\ 36.02 & 24.17 & 37.84 & 31.83 & 25.59 & 37.36 & 25.69 & 34.74 & 25.46 & -278.71 \end{bmatrix}$$

Table 3.1 provides a summary of the Cramér-von Mises test statistics when comparing cumulative probability values obtained by simulating the true process with generator $\mathbf{Q}$ with those of numerical Laplace transform inversion obtained using

$\hat{\mathbf{Q}}_T$. Three observation periods were considered in this experiment, and for the remaining lifetime tests, we fixed $\xi_0 = \mathbb{E}[T(x)]$.

Table 3.1: Cramér-von Mises test statistics for Case I (κ^*=0.461, $\alpha = 0.05$).

Distribution	Run Length	$K = 2$	$K = 5$	$K = 10$
$\hat{F}(x,t)$ (m=48)	$T = 100$	7.013E-03	4.233E-02	9.477E-03
	$T = 500$	1.649E-03	7.895E-03	2.130E-03
	$T = 5000$	9.941E-05	9.019E-05	1.986E-06
$1 - \hat{R}(x,t\|\xi_0)$ (m=50)	$T = 100$	5.511E-03	3.821E-02	7.455E-03
	$T = 500$	1.223E-03	6.553E-03	2.113E-03
	$T = 5000$	5.544E-05	8.092E-05	2.576E-06

It is noted that, in all eighteen experiments, we fail to reject the null hypothesis that the two cumulative distribution functions are equivalent. Thus, if the environment process can be partitioned into K distinct and observable states with known degradation rates, our approach provides a viable approximation procedure that does not require lifetime observations; it requires only a count of environment transitions. Case II compares the approximated distribution resulting from the estimation of the degradation rates and the generator matrix given a known number of environment states with the true distribution resulting from the known parameters.

Case II. Observable Degradation

For this scenario, we first simulate linear degradation paths generated via a known environment process with $K = 3$ distinct states and known degradation rates $\mathcal{D} = \{r(1), r(2), \ldots, r(K)\}$. Five hundred degradation sample paths, similar to the five shown in Figure 3.7, were generated for this experiment.

Setting $t_0 \equiv 0$, we observe the level of degradation at M equally spaced points in time such that $t_M = 20$. That is, the degradation process is observed up to time 20. Figure 3.8 on the left depicts a sample of observations for five degradation sample paths. Additionally, Figure 3.8 shows how these observations were used to form new piecewise, linear degradation sample paths.

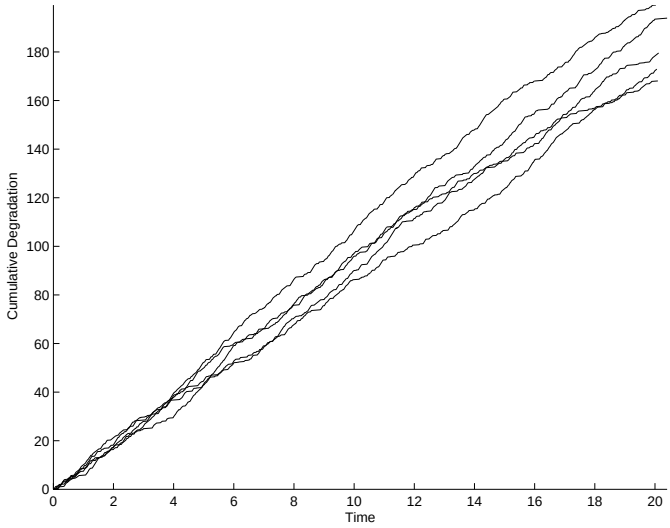

Figure 3.7: A sample of five linear degradation paths.

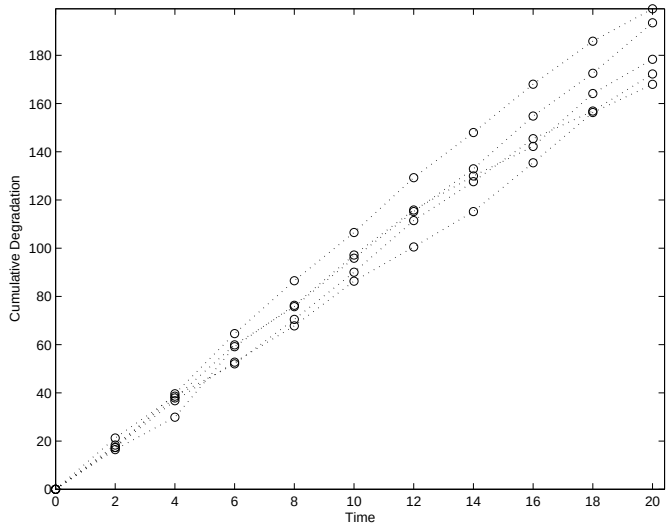

Figure 3.8: Piecewise approximations, $M = 10$, to the linear degradation paths.

Combining Figures 3.7 and Figure 3.8 visually shows how we are using the simulated data to form new sample paths for estimation purposes. This combination is shown in Figure 3.9.

Before forming a new sample path from each of the 500 generated degradation paths, we visually ensured that the true lifetime distribution, as determined from numerical inversion using the known generator matrix and degradation rates, was a

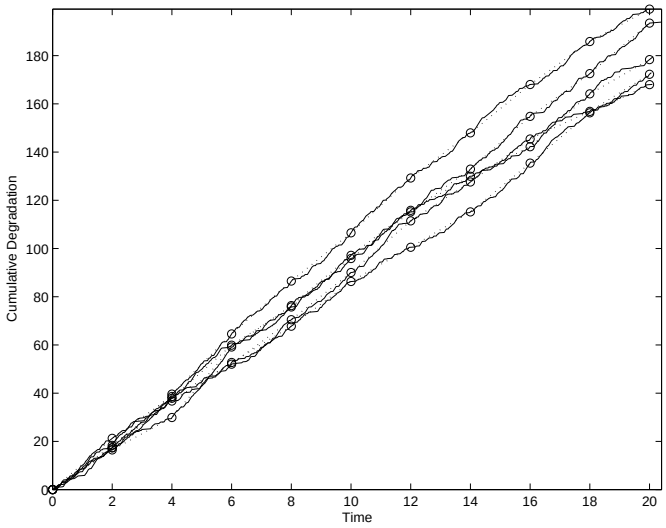

Figure 3.9: Simulated versus approximated degradation sample paths.

close approximation to the empirical lifetime distribution formed from the 500 time observations of the simulated degradation paths at a degradation value of 152.6. After examining the results in Figure 3.10, we decided to continue our estimation technique.

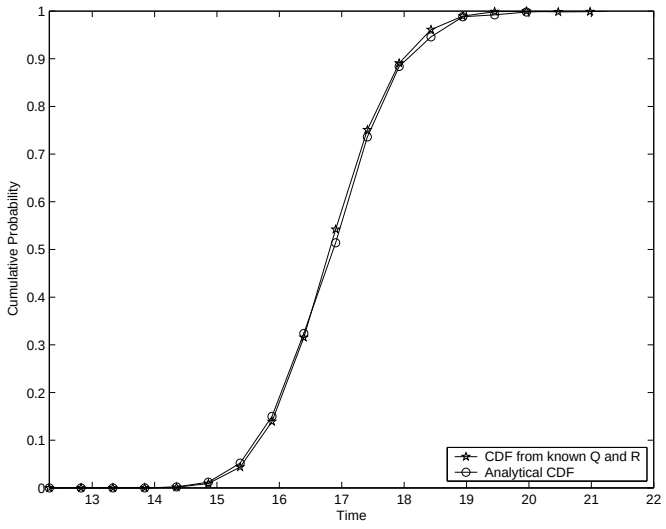

Figure 3.10: Analytical CDF versus Simulated CDF

Following this confirmation, we utilize these piecewise linear approximations to obtain the lifetime and remaining lifetime distribution. In each scenario, we observe the system up to time $t_M = 20$ but vary the inter-observation time. In particular, we consider $M = 20$, 100, and 200 observations on $[0, 20]$, respectively. The assumed unit failure threshold was fixed at $x = 152.6$ units. The same 500 simulated degradation sample paths were used for each value of M to estimate a new generator matrix and new degradation rates at times 4.0, 8.0, and 12.0, respectively. The generator matrices are shown in Table 3.2 and the degradation rates are shown in Table 3.3.

Table 3.2: Estimated generator matrices for each M at three times.

M	$T = 4$	$T = 8$	$T = 12$
20	$\begin{bmatrix} -0.81 & 0.53 & 0.29 \\ 0.38 & -0.71 & 0.33 \\ 0.28 & 0.55 & -0.84 \end{bmatrix}$	$\begin{bmatrix} -0.75 & 0.48 & 0.27 \\ 0.33 & -0.62 & 0.28 \\ 0.27 & 0.51 & -0.78 \end{bmatrix}$	$\begin{bmatrix} -0.74 & 0.48 & 0.26 \\ 0.31 & -0.60 & 0.29 \\ 0.26 & 0.50 & -0.76 \end{bmatrix}$
50	$\begin{bmatrix} -1.81 & 0.64 & 1.17 \\ 0.68 & -1.91 & 1.23 \\ 0.83 & 0.68 & -1.52 \end{bmatrix}$	$\begin{bmatrix} -1.76 & 0.58 & 1.18 \\ 0.67 & -1.85 & 1.17 \\ 0.79 & 0.67 & -1.46 \end{bmatrix}$	$\begin{bmatrix} -1.78 & 0.63 & 1.14 \\ 0.63 & -1.79 & 1.16 \\ 0.74 & 0.70 & -1.45 \end{bmatrix}$
100	$\begin{bmatrix} -3.30 & 2.08 & 1.23 \\ 1.55 & -2.98 & 1.42 \\ 1.29 & 2.15 & -3.44 \end{bmatrix}$	$\begin{bmatrix} -3.30 & 2.12 & 1.18 \\ 1.54 & -2.89 & 1.35 \\ 1.26 & 2.19 & -3.46 \end{bmatrix}$	$\begin{bmatrix} -3.35 & 2.15 & 1.20 \\ 1.52 & -2.89 & 1.37 \\ 1.25 & 2.18 & -3.43 \end{bmatrix}$
200	$\begin{bmatrix} -5.88 & 3.86 & 2.02 \\ 3.21 & -5.86 & 2.65 \\ 2.30 & 4.41 & -6.71 \end{bmatrix}$	$\begin{bmatrix} -5.84 & 3.85 & 1.99 \\ 3.22 & -5.85 & 2.63 \\ 2.38 & 4.35 & -6.73 \end{bmatrix}$	$\begin{bmatrix} -5.86 & 3.88 & 1.99 \\ 3.21 & -5.83 & 2.62 \\ 2.37 & 4.32 & -6.69 \end{bmatrix}$
500	$\begin{bmatrix} -10.17 & 6.71 & 3.46 \\ 7.42 & -13.18 & 5.76 \\ 3.65 & 9.91 & -13.56 \end{bmatrix}$	$\begin{bmatrix} -10.07 & 6.73 & 3.34 \\ 7.35 & -13.10 & 5.75 \\ 3.60 & 9.86 & -13.46 \end{bmatrix}$	$\begin{bmatrix} -10.00 & 6.70 & 3.30 \\ 7.35 & -13.12 & 5.77 \\ 3.60 & 9.87 & -13.47 \end{bmatrix}$
Actual		$\begin{bmatrix} -16.10 & 12.04 & 4.06 \\ 16.84 & -26.52 & 9.69 \\ 5.98 & 21.76 & -27.75 \end{bmatrix}$	

Table 3.3: Estimated degradation rates for each M at three times.

Samples (M)	Run Length	$\hat{r}(1)$	$\hat{r}(2)$	$\hat{r}(3)$
20	$T = 4$	0.0441	0.0600	0.0765
	$T = 8$	0.0443	0.0599	0.0762
	$T = 12$	0.0441	0.0597	0.0759
50	$T = 4$	0.0361	0.0847	0.0604
	$T = 8$	0.0366	0.0848	0.0604
	$T = 12$	0.0361	0.0838	0.0596
100	$T = 4$	0.0285	0.0602	0.0924
	$T = 8$	0.0287	0.0605	0.0928
	$T = 12$	0.0284	0.0601	0.0923
200	$T = 4$	0.0207	0.0625	0.1039
	$T = 8$	0.0211	0.0629	0.1040
	$T = 12$	0.0210	0.0628	0.1038
500	$T = 4$	0.0133	0.0658	0.1179
	$T = 8$	0.0134	0.0657	0.1177
	$T = 12$	0.0135	0.0659	0.1177
Actual		0.0066	0.0655	0.1311

It is very interesting to note how both the generator matrices and the degradation rates slightly vary from times $T = 4.0$, $T = 8.0$, and $T = 12.0$ for each value of M. Additionally, we notice that both the generator matrices and degradation rates are converging to the true generator matrix and degradation rates as the number of observations over the time interval increases. Of particular interest is how the second degradation rate is very close to the exact rate while the first and third rate are not as close. This can be explained by the K-means clustering technique. The interior degradation rates are better estimated by the centroid of each cluster while the exterior points are not. For $r(1)$ and its estimate $\hat{r}_i(1)$, $\hat{r}_i(1) \geq r(1)$ and for $r(K)$ and its estimate $\hat{r}_j(K)$, $\hat{r}_j(K) \leq r(K)$, for all i, j. However, as we will demonstrate, this difference in the outer degradation rates does not significantly affect the overall results. Given the estimates at each M, the next step is to compare the distributions obtained from numerical inversion of the lifetime distribution. The estimated generator matrices and degradation rates at times $T = 4.0$, $T = 8.0$, and $T = 12.0$, for each value of M are examined and are shown in Figures 3.11, 3.12 and 3.13.

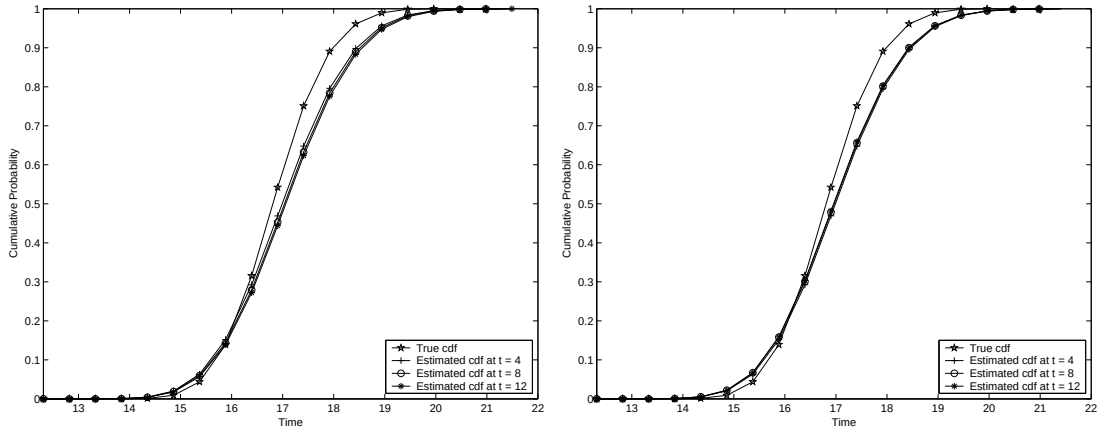

Figure 3.11: CDF comparison with $t_{20} = 20$ (left) and CDF with $t_{50} = 20$ (right).

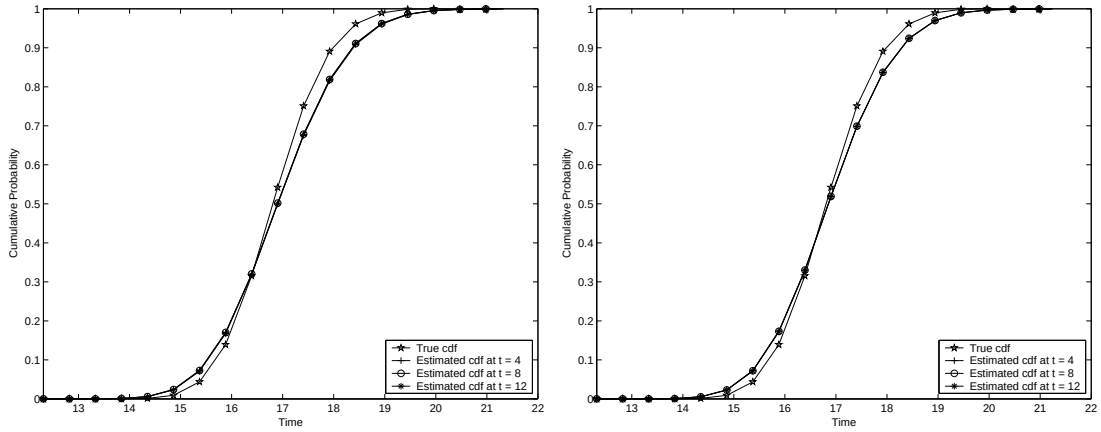

Figure 3.12: CDF comparisons with $t_{100} = 20$ (left) and CDF with $t_{200} = 20$ (right).

While the lifetime distributions in these figures appear close, we checked the equality of the lifetime distributions, as well as the remaining lifetime distributions, by comparing the approximated distribution with the true distribution via the Cramér-von Mises goodness-of-fit test at $m = 20$ using Equation (3.88). The results of this experiment, summarized in Table 3.4, indicate that we would fail to reject the null hypothesis that the approximated and true distributions are the same for the 18 distribution comparisons (nine lifetime and nine remaining lifetime).

Thus, by using observations of unit degradation up to a fixed point in time, we conclude that we are able to adequately approximate the lifetime and remaining

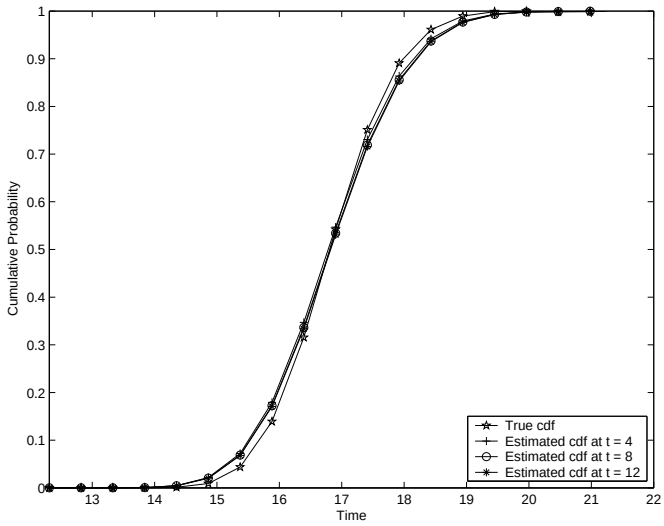

Figure 3.13: CDF comparison with $t_{500} = 20$.

Table 3.4: Cramér-von Mises statistic for Models 1-4 (κ^*=0.461 at $\alpha = 0.05$).

M	Interval	$\hat{F}(x,t)$	ξ_0	$1-\hat{R}(x,t\|\xi_0)$
20	[0.0,4.0]	0.0157	4.0	0.0189
	[0.0,8.0]	0.0210	8.0	0.1213
	[0.0,12.0]	0.0247	12.0	0.1633
100	[0.0,4.0]	0.0095	4.0	0.0197
	[0.0,8.0]	0.0088	8.0	0.0476
	[0.0,12.0]	0.0087	12.0	0.0639
200	[0.0,4.0]	0.0053	4.0	0.0149
	[0.0,8.0]	0.0051	8.0	0.0336
	[0.0,12.0]	0.0052	12.0	0.0266

lifetime distribution of the component for these simulated degradation paths via estimation of the degradation rates and generator matrix. We now turn our attention to Case III where we examine the feasibility of obtaining a good approximation of the lifetime and remaining lifetime distributions via estimation of the number of environment states, the degradation rates, and the generator matrix.

Case III. Observable Degradation

To illustrate our method to determine the number of environment states $\hat{K}$, we began with models having a known number of environment states. We noticed our method to select $\hat{K}$ worked well in many scenarios, but some scenarios designated fewer states than the known number of environment states. Four examples, two modelled with a 5-state environment (Models 1 and 2) and two modelled with a 10-state environment (Models 3 and 4), were examined. Using Equation (3.83), for $K = 2, \ldots, 15$, and choosing the first local maximum, we can find $\hat{K}$. Table 3.5 contains the F_K values for $K = 2, \ldots, 15$, for all four models where $\hat{K}$ is underlined.

Table 3.5: Values of $F_K(\times 10^5)$

K	Model 1	Model 2	Model 3	Model 4
2	1.6549729	0.8181652	0.2850496	1.2135811
3	3.0908480	<u>1.2615370</u>	0.4007699	1.4698324
4	<u>4.0673686</u>	1.1724026	0.4211020	2.0085264
5	3.8208568	1.5243076	0.6602448	<u>2.6493771</u>
6	3.2499389	2.2097990	0.6982035	2.0623002
7	6.6786389	2.2156051	0.9551702	3.6891118
8	7.2785629	2.8230510	<u>1.1446319</u>	4.0491842
9	10.762887	3.2351094	1.0479320	5.0331320
10	7.3831984	2.5987248	0.6973468	4.6961658
11	5.8765496	3.7692757	1.3063962	6.3385019
12	9.9207365	4.3081416	1.8050071	8.0903999
13	8.5587210	2.7749692	0.8572278	5.2739689
14	15.161074	5.2694757	0.8588647	10.272132
15	12.455254	5.6134761	1.3868170	9.9281135

As shown in Table 3.5, the known 5-state environments have $\hat{K} = 4$ and $\hat{K} = 3$ for Model 1 and Model 2 respectively while the known 10-state environments have $\hat{K} = 8$ and $\hat{K} = 5$ for Model 3 and Model 4 respectively. To determine if each $\hat{K}$ adequately estimates the known K value, we estimated the degradation rates and generator matrix associated with each of the four models. Additionally, we examined $\hat{K} + 1$ for each model. We then determined the approximated and true lifetime distributions using Equation (3.51) and compared these distributions using

the Cramér-von Mises test. The graphs for Model 2 and Model 4 are shown in 3.14 and the summarized results for all models are contained in Table 3.6. Thus, using the Cramér-von Mises test, we fail to reject the null hypothesis that the approximated and true lifetime distributions are the same.

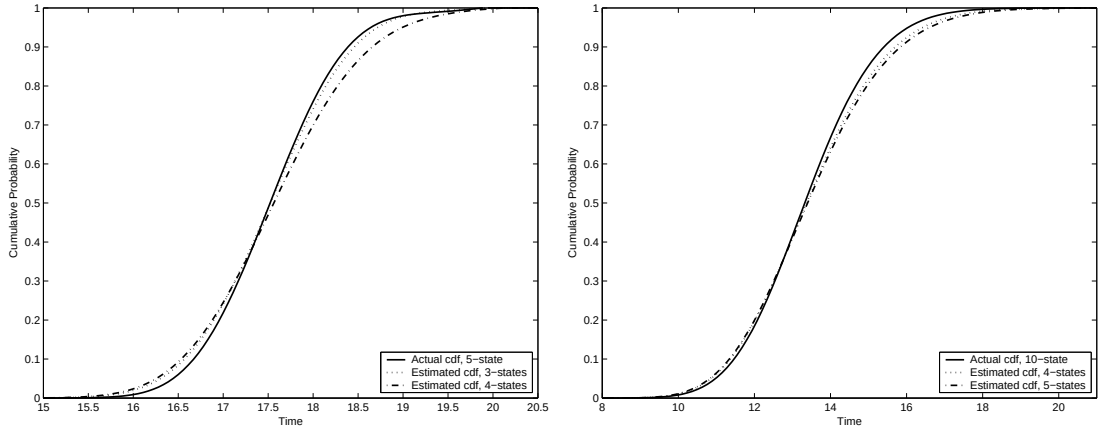

Figure 3.14: Distributions with known 5-state (left) and 10-state (right).

Table 3.6: Cramér-von Mises statistic for Models 1-4 (κ^*=0.461 at $\alpha = 0.05$).

Model	K	$\hat{K}$	κ_2
1	5	4	2.3056E-02
	5	5	3.2317E-02
2	5	3	8.2275E-03
	5	4	4.7578E-02
3	10	8	3.5704E-02
	10	9	3.5397E-02
4	10	5	2.0722E-02
	10	6	2.7385E-02

To conclude this case, we chose Models 1 and 3 to examine the remaining lifetime distribution because these two models most closely estimate the actual number of states. Those results, shown in Table 3.7 also indicate that we fail to reject the null hypothesis that the approximated and true remaining lifetime distribution are the same.

Table 3.7: Cramér-von Mises statistic for Model 1 and Model 3 (κ^*=0.461 at $\alpha = 0.05$).

Model	K	$\hat{K}$	ξ_0	κ_2
1	5	4	17.5311	1.0179E-01
	5	5	17.5311	1.0733E-01
3	10	8	13.4249	1.6783E-01
	10	9	13.4249	1.4709E-01

Case IV. Observable Degradation

Finally, we illustrate the implementation of Equation (3.51) and the associated estimation techniques using real degradation data. Virkler, et al. [90] produced this degradation data by tracking the number of load cycles required to grow a crack in 2024-T3 aluminum alloy under specified stress. Degradation paths were obtained for 68 test specimens and are shown in Figure 3.6. For a more thorough treatment of fatigue crack dynamics, the reader is referred to the paper by Ray and Tangirala [76] and the references contained therein.

We begin our procedure by letting $X(w)$ denote the length of the crack at time w and assuming that the (linear) rate at which the crack grows is subject to its random environment (applied stress, ambient conditions, and other factors). It is assumed that these environmental factors can be characterized by $\{Z(w) : w \geq 0\}$, a homogeneous Markov process on a finite-state sample space $S = \{1, 2, \ldots K\}$. We observe all 68 sample paths to estimate the generator matrix, $\mathbf{Q}$, for the $\mathcal{Z}$ process. In the experiment that follows, the component is said to fail whenever the crack length exceeds a value of 45 mm. We estimate the off-diagonal generator matrix values by

$$q_{i,j} \approx \hat{q}^{(68)} = \frac{N^{(68)}(i,j)}{H^{(68)}(i)} \tag{3.89}$$

as defined in Equation (3.75). As a quality check, we hypothesize that the estimated cumulative probability values will converge to the true cumulative probability values as the observation period approaches the failure time. It should be noted, however, that our 'true' cumulative probability values are obtained via the empirical cumula-

tive probabilities for the 68 degradation paths at $x = 45$ mm. We assume that these 68 observations provide a 'close' approximation to the underlying true distribution.

We designate three points in the data, specifically 25 mm, 35 mm, and 41 mm, which correspond to approximately 55%, 77%, and 91% of the lifetime. Setting $K = 3$, we initially demonstrate that our hypothesis is correct after estimating the degradation rates and the generator matrix. Using Equation (3.51) to approximate the distribution, we can see in Figure 3.15 that we are approaching the empirical distribution for $x = 45$ mm. Additionally, using Equations (3.65) and (3.69) at $n = 1$, we compute the mean lifetime required to reach a fixed threshold value x as well as the mean residual lifetime. For each case, we estimate a new generator matrix and degradation rates. In the actual data sets, we have that the initial measured crack length is approximately 9.0 mm (i.e., $X(0) \approx 9.00$ mm). Table 3.8 provides the comparison of the level-crossing times over the 68 degradation paths. Using Figure 3.15 and Table 3.8, we feel confident that we can adequately estimate the parameters of the lifetime distribution and proceed with the entirety of our approach.

Table 3.8: Mean lifetime and remaining lifetimes ($\times 10^5$), ξ_0 = actual $m^{(1)}(x)$.

	$m^{(1)}(x)$		$m^{(1)}(x\|\xi_0)$	
x (mm)	Actual	Model	Actual	Model
10	0.318574	0.325974	0.349943	0.347018
15	1.183405	1.150109	1.268688	1.293640
20	1.638693	1.530864	1.738733	1.771963
25	1.938014	1.800568	2.042441	2.113597
30	2.158575	2.021006	2.286812	2.368369
35	2.326253	2.116553	2.458462	2.508772
41	2.464467	2.298747	2.611653	2.633815

Applying our methodology, we begin by using F_K to estimate the number of environment states, $\hat{K}$ at the three designated points. As seen in Table 3.9, we select two clusters at 55% of the lifetime, nine clusters at 77% of the lifetime (due to the sharp increase), and twelve clusters at 91% of the lifetime. Since 91% of the lifetime

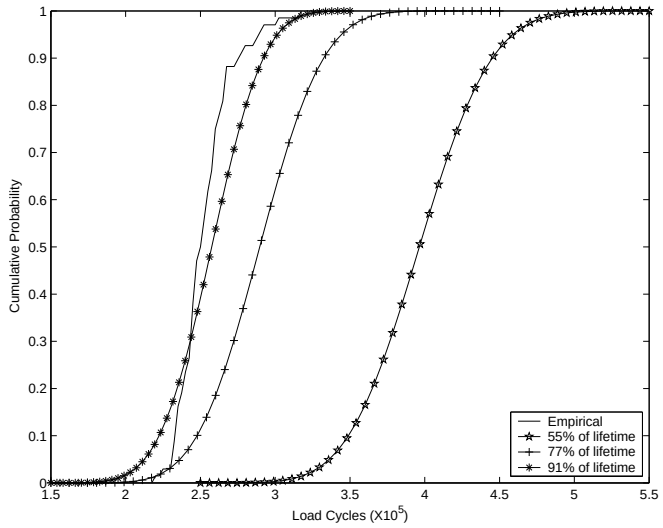

Figure 3.15: Comparison of cumulative probability values.

appears to offer the closest match to the empirical distribution in Figure 3.15, we focus our attention on this set of data.

Table 3.9: Values of $F_K(\times 10^4)$.

K	55%	77%	91%
2	0.845912	1.716694	1.934667
3	0.770946	2.045060	2.448228
4	0.700849	2.267842	2.866000
5	0.637918	2.412888	3.248069
6	1.789474	2.506080	3.620849
7	1.984733	2.627177	3.934574
8	2.230180	2.658314	4.208318
9	2.570793	2.669967	4.374201
10	2.785035	4.566412	4.448147
11	3.017830	4.974103	4.744162
12	2.993362	5.272737	4.830633
13	3.714449	5.683212	4.796518
14	4.167025	5.850034	4.800008
15	4.292031	6.608150	4.745232

Estimates of the generator matrix and degradation rates were constructed for $\hat{K} = 2, 3, 4, 9$, and 12 using 91% of the data. The resulting distributions for each

scenario were compared to the empirical distribution at a fixed number of points ($m = 65$) using the Cramér-von Mises test and are summarized in Table 3.10.

Table 3.10: Cramér-von Mises test statistic ($\kappa^* = 0.461$ at $\alpha = 0.05$).

$\hat{K}$	$\hat{F}(x,t)$	ξ_0	$1 - \hat{R}(x,t\|\xi_0)$
2	1.398488E-01	2.534	6.124815E-01
3	8.671149E-02	2.534	1.395285E-01
4	4.525246E-01	2.534	4.383202E-01
9	7.657227E-01	2.534	2.235864E-00
12	8.723972E-01	2.534	1.667689E-00

For the lifetime distribution, we fail to reject the null hypothesis that the distributions are equal in the Cramér-von Mises test for $\hat{K} = 2, 3$, and 4. For the residual lifetime distributions, we fail to reject the null hypothesis that the distributions are equal in the Cramér-von Mises test for $\hat{K} = 3$ and 4. At this point, our estimate for the number of environment states is suspect because the F_K-value in Table 3.9 for 91% of the data indicates that twelve states is the best estimate for K. However, as shown in Table 3.10 for $\hat{K} = 12$, we reject the null hypothesis that the distributions are equivalent. A possible reason for the lack of fit may be that we require additional data to adequately estimate the elements of the generator matrix and associated degradation rates. However, this issure cannot be explored as additional data is not available. Other possible reasons for the poor fit at $\hat{K} = 12$ may be that the state holding times are not exponentially distributed or that the data might be nonstationary.

Results for this example suggest that a smaller representation may exist that allows a better approximation of the lifetime distribution. Table 3.10 indicates that $\hat{K} = 3$ or 4 adequately approximates the empirical distribution. However, additional research is required to determine if a smaller representation exists.

This chapter has provided analytical results for the lifetime distribution of the time required for a single-unit system to reach a prespecified cumulative degradation threshold operating in a nonhomogeneous and homogeneous Markovian environment.

For the case of a homogeneous Markov environment, given a sensor that tracks and relays information about the known environment or the real-time degradation level of a component, our technique provides a capability to obtain the remaining lifetime distribution of that component and provide failure time probabilities required for systems prognosis.

While, this methodology is a contribution to degradation-based reliability [45], it is quite restrictive. By assuming a Markovian environment, our results are limited to environments that spend an exponential amount of time in each state. In reality, this assumption will not be valid in many cases which motivates the need to relax this assumption. The next chapter explores a semi-Markov environment which allows the amount of time spent in each environment state to be generally distributed.

IV. Homogeneous Semi-Markov Environments

Chapter III provided analytical results and numerical procedures using observable environment and observable degradation data to obtain the remaining lifetime distribution required for systems prognosis. These results assumed the system was operating in a Markovian environment which required the amount of time spent in each state of the environment to be exponentially distributed. In reality, however, this assumption may not hold. In this chapter, we relax the assumption of exponential holding times by assuming the system operates in a semi-Markovian environment. This environment allows transitions from one state to another in the same manner as the Markovian environment, but does not require that the amount of time spent in each state be exponentially distributed. This procedure substantially improves the ability to model a variety of real-world environments as compared to the procedures of Chapter III.

In this chapter, we provide a brief review of semi-Markov processes (SMP) by highlighting some of their fundamental definitions and properties. Following this review, in Section 4.2 we present the main results for the lifetime distribution of a single-unit system that continuously degrades in an environment modelled as a homogeneous SMP. In Section 4.3, we provide three illustrative examples that demonstrate the difficulties associated with numerical implementation of this lifetime distribution result, but then establish in Chapter V that these difficulties can be overcome to ultimately extend the utility of the procedures developed in Chapter III.

4.1 Review of Semi-Markov Processes

The time homogeneous semi-Markov process (SMP) is a generalization of the homogeneous continuous-time Markov chain (CTMC). Both processes can have an infinite number of states and also have embedded at transition epochs a discrete-time

Markov chain (DTMC), $\{Y_n : n \geq 0\}$ with transition probabilities

$$p_{i,j} = P\{Y_{n+1} = j | Y_n = i\}. \tag{4.1}$$

Let $\mathcal{Z} \equiv \{Z(w) : w \geq 0\}$ be a continuous-time stochastic process. As seen in Figure 4.1, define τ_n as the nth jump epoch and let $Y_n \equiv Z(\tau_n^+)$ be the state of the environment just after the nth transition.

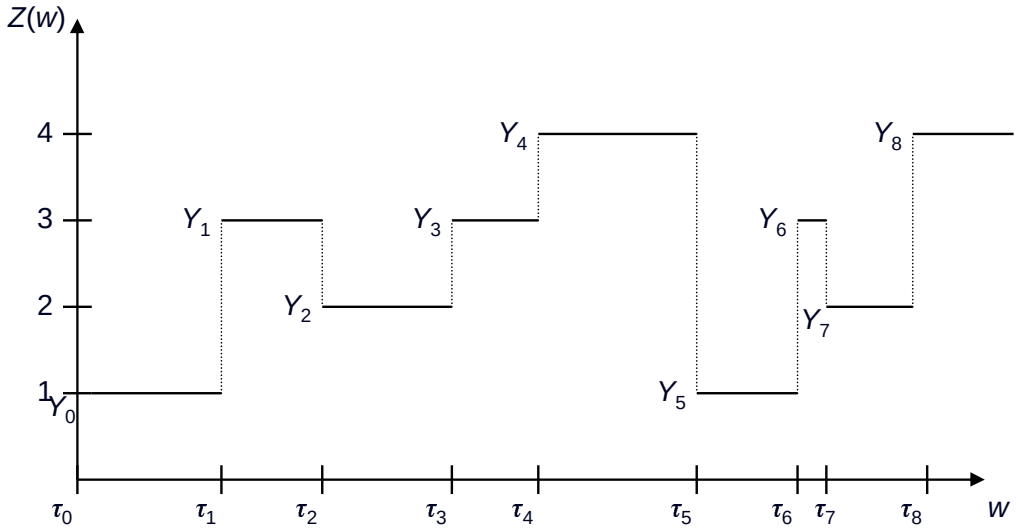

Figure 4.1: Sample path of a semi-Markov process.

Recall that a sequence of bivariate random variables, $\{(Y_n, \tau_n) : n \geq 0\}$ is a Markov renewal sequence if, for all $n \geq 0$,

$$\begin{aligned} \Psi_{i,j}(w) &= P\{\tau_{n+1} - \tau_n \leq w, Y_{n+1} = j | \tau_n, Y_n = i, \ldots, \tau_0, Y_0\} \\ &= P\{\tau_{n+1} - \tau_n \leq w, Y_1 = j | Y_0 = i\}. \end{aligned} \tag{4.2}$$

where $\tau_0 = 0, \tau_{n+1} \geq \tau_n$, and $Y_n \in S$. If the stochastic process has piecewise constant, right continuous sample paths and $\{(Y_n, \tau_n) : n \geq 0\}$ is a Markov renewal sequence, then $\mathcal{Z}$ is a semi-Markov process [50] with semi-Markov kernel $\mathbf{\Psi}(w) = [\Psi_{i,j}(w)]$. A

homogeneous semi-Markov process has

$$\Psi_{i,j}(w) = \{\tau_1 \leq w, Y_1 = j | Y_0 = i\}. \tag{4.3}$$

Holding times, or sojourn times, are defined by $\tau_n - \tau_{n-1}, n \geq 1$ and the holding time distributions are

$$H_i(w) = \sum_{j=0}^{\infty} \Psi_{i,j}(w). \tag{4.4}$$

This short review of a semi-Markov process provides the necessary definitions that lead to the development and derivation of the lifetime distribution and the illustrative examples in the remaining sections of this chapter.

4.2 Lifetime Distribution Results

If the environment $\mathcal{Z}$ is modelled as a semi-Markov process, the Markov property still holds with respect to the embedded DTMC but not with respect to the state holding times which are not exponentially distributed in general. Thus, we cannot apply the standard techniques employed in Section 3.2 to obtain the lifetime and remaining lifetime distribution.

For a system subject to a semi-Markov environment, we derive the lifetime distribution via a bivariate distribution based upon Equation (3.6) similar to that shown in Kulkarni, *et al.*, [52]. Let the rate of degradation of the system at time $w > 0$ be governed by the random environment modelled as a homogeneous semi-Markov process (SMP). Specifically, the time spent in any particular state of the environment is generally distributed. Recall that $\mathcal{Z} = \{Z(w) : w \geq 0\}$ represents a stochastic process with sample space $S = \{1, ..., K\}$, where K is a positive integer, and the random variable $Z(w)$ represents the state of the random environment at time w. Associated with each state of the environment is a degradation rate where

the collection of all degradation rates is again given by $\mathcal{D} = \{r(1), \ldots, r(K)\}$. If $Z(w) = i \in S$, then $r(Z(w)) = r(i)$. Additionally, the environment jumps from state $i \in S$ to state $j \in S$ according to the one-step embedded transition probability matrix $\mathbf{P} \equiv [p_{i,j}]$, where $p_{i,j} \equiv P\{Y_1 = j | Y_0 = i\}$.

Let

$$\psi_{i,j}(w) = \frac{d\Psi_{i,j}(w)}{dw}$$

where $\boldsymbol{\psi}(w) \equiv [\psi_{i,j}(w)]$. Define the conditional lifetime distribution

$$F_i(x, w) \equiv P\{T(x) \leq w | Z(0) = i\}$$

and define the Laplace-Stieltjes transform (LST) of $F_i(x, w)$ with respect to w as

$$\widetilde{F}_i(x, s) \equiv \mathbb{E}(e^{-sT(x)} | Z(0) = i)$$

so that

$$\widetilde{F}(x, s) \quad = \quad \sum_{i=1}^{n} \widetilde{F}_i(s, x) P\{Z(0) = i\}$$

denotes the LST of the unconditional distribution $F(x, w) \equiv P\{T(x) \leq w\}$ with respect to w. Additionally, let the Laplace transform of $\widetilde{F}(x, s)$ with respect to x be $\widetilde{F}^*(u, s)$. Theorem 7 provides the main result of this chapter.

Theorem 7. *If the environment process, $\mathcal{Z}$, with finite-state space S, is a homogeneous semi-Markov process, then*

$$\widetilde{F}^*(u, s) = \boldsymbol{\alpha}[\mathbf{I} - \boldsymbol{\psi}^*(u, s)]^{-1}[\mathbf{A}(u, s) - \mathbf{H}^*(u, s)]\mathbf{r} \tag{4.5}$$

where $\boldsymbol{\alpha}$ *is the initial probability distribution vector of the* $\mathcal{Z}$ *process,* $\mathbf{I}$ *is a* $K \times K$ *identity matrix, the matrix*

$$\boldsymbol{\psi}^*(u,s) = \begin{bmatrix} \psi^*_{1,1}(s+r(1)u) & \psi^*_{1,2}(s+r(1)u) & \cdots & \psi^*_{1,K}(s+r(1)u) \\ \psi^*_{2,1}(s+r(2)u) & \psi^*_{2,2}(s+r(2)u) & \cdots & \psi^*_{2,K}(s+r(2)u) \\ \vdots & \vdots & \ddots & \vdots \\ \psi^*_{K,1}(s+r(K)u) & \psi^*_{K,2}(s+r(K)u) & \cdots & \psi^*_{K,K}(s+r(K)u) \end{bmatrix},$$

$\mathbf{A}(u,s) = diag[1/(s+r(1)u), 1/(s+r(2)u), \ldots, 1/(s+r(K)u)]$, $\mathbf{H}^*(u,s) = diag[H^*_1(s+r(1)u), H^*_2(s+r(2)u), \ldots, H^*_K(s+r(K)u)]$, *and*

$$\mathbf{r} \equiv \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(K) \end{bmatrix}.$$

The result will be proved by a standard Markov renewal argument.

Proof.

$$\mathbb{E}(e^{-sT(x)}|\tau_1 = h, Z(0) = i) = \begin{cases} e^{-sx/r(i)} & \text{if } h \geq \frac{x}{r(i)}, \quad \text{(Case 1)} \\ e^{-sh} \sum_j p_{i,j} \widetilde{F}_j(x - r(i)h, s) & \text{if } h < \frac{x}{r(i)}, \quad \text{(Case 2).} \end{cases}$$

For Case 1, the LST of the unit lifetime is $\exp(-sx/r(i))$ because the system first transitions after time $x/r(i)$. In Case 2, the system first transitions before time $x/r(i)$ and hence, must account for the remaining time between h and $x/r(i)$. Un-

conditioning with respect to the initial state, we obtain

$$\begin{aligned}
\widetilde{F}_i(x,s) &= \int_0^\infty \mathbb{E}(e^{-sT(x)}|\tau_1 = h, Z(0) = i) dH_i(h) \\
&= \int_0^{\frac{x}{r(i)}} e^{-sh} \sum_j p_{i,j} \widetilde{F}_j(x - r(i)h, s) dH_i(h) + \int_{\frac{x}{r(i)}}^\infty e^{-sx/r(i)} dH_i(h) \\
&= \sum_j p_{i,j} \int_0^{\frac{x}{r(i)}} e^{-sh} \widetilde{F}_j(x - r(i)h, s) dH_i(h) \\
&\quad + e^{-sx/r(i)} \left(1 - H_i\left(\frac{x}{r(i)}\right)\right). \qquad (4.6)
\end{aligned}$$

If $Z(w)$ depends only on the previous state and not on τ_1, then

$$\begin{aligned}
\Psi_{i,j}(w) &= P\{\tau_1 \leq w, Z(\tau_1^+) = j | Z(0) = i\} \\
&= P\{Z(\tau_1^+) = j | Z(0) = i\} P\{\tau_1 \leq w | Z(0) = i\} \\
&= p_{i,j} H_i(w)
\end{aligned}$$

which allows Equation (4.6) to also be written as

$$\widetilde{F}_i(x,s) = \sum_j \int_0^{\frac{x}{r(i)}} e^{-sh} \widetilde{F}_j(x - r(i)h, s) d\Psi_{i,j}(h) + e^{-sx/r(i)} \left(1 - H_i\left(\frac{x}{r(i)}\right)\right) \qquad (4.7)$$

Taking the Laplace transform of Equation (4.6) with respect to x and doing a change of variable, where we designate $y = r(i)h$ and $dy = r(i)dh$, results in

$$\widetilde{F}_i^*(u,s) = \int_0^\infty e^{-ux}\widetilde{F}_i(x,s)dx$$

$$= \int_0^\infty e^{-ux}\left[\sum_j p_{i,j}\int_0^{\frac{x}{r(i)}} e^{-sh}\widetilde{F}_j(x-r(i)h,s)dH_i(h) + e^{-sx/r(i)}\left(1-H_i\left(\frac{x}{r(i)}\right)\right)\right]dx$$

$$= \int_0^\infty e^{-ux}\left[\sum_j p_{i,j}\int_0^{\frac{x}{r(i)}} e^{-sh}\widetilde{F}_j(x-r(i)h,s)h_i(h)dh + e^{-sx/r(i)}\left(1-H_i\left(\frac{x}{r(i)}\right)\right)\right]dx$$

$$= \sum_j p_{i,j}\int_0^\infty e^{-ux}\int_0^x \widetilde{F}_j(x-y,s)e^{-sy/r(i)}h_i\left(\frac{y}{r(i)}\right)\frac{dy}{r(i)}dx + \int_0^\infty e^{-ux}e^{-sx/r(i)}dx$$
$$- \int_0^\infty e^{-ux}e^{-sx/r(i)}H_i\left(\frac{x}{r(i)}\right)dx$$

$$= \sum_j \frac{p_{i,j}}{r(i)}\int_0^\infty e^{-ux}\int_0^x \widetilde{F}_j(x-y,s)e^{-sy/r(i)}h_i\left(\frac{y}{r(i)}\right)dydx + \frac{1}{\frac{s}{r(i)}+u}$$
$$- \int_0^\infty e^{-ux}e^{-sx/r(i)}H_i\left(\frac{x}{r(i)}\right)dx$$

$$= \sum_j \frac{p_{i,j}}{r(i)}\int_0^\infty e^{-ux}\int_0^x \widetilde{F}_j(x-y,s)e^{-sy/r(i)}h_i\left(\frac{y}{r(i)}\right)dydx + \frac{r(i)}{s+r(i)u}$$
$$- \int_0^\infty e^{-ux}e^{-sx/r(i)}H_i\left(\frac{x}{r(i)}\right)dx. \quad (4.8)$$

Oberhettinger and Badii [67] provided the Laplace transform

$$H_i^*(r(i)u-s) = \int_0^\infty e^{-ux}\frac{e^{sx/r(i)}H_i\left(\frac{x}{r(i)}\right)}{r(i)}dx. \quad (4.9)$$

Using this transform for the last integral in Equation (4.8), we obtain

$$\begin{aligned}\widetilde{F}_i^*(u,s) &= \sum_j \frac{p_{i,j}}{r(i)} \int_0^\infty e^{-ux} \int_0^x \widetilde{F}_j(x-y,s) e^{-sy/r(i)} h_i\left(\frac{y}{r(i)}\right) dydx \\ &\quad + \frac{r(i)}{s+r(i)u} - r(i)H_i^*(s+r(i)u). \end{aligned} \tag{4.10}$$

This result may also be obtained by performing a change of variable with $z = x/r(i)$ and $r(i)dz = dx$ and applying the Laplace transform damping theorem where $\mathcal{L}\{e^{-at}f(t)\} = f^*(s+a)$. To reduce the remaining double integral, we note that

$$H(y,s) = \int_0^x \widetilde{F}_j\left(x-y,s\right) g(y,s) dy \Longrightarrow H^*(u,s) = \widetilde{F}_j^*\left(u,s\right) g^*(u,s).$$

This implies that $g(y,s) = \exp(-sy/r(i))h_i(y/r(i))$ where Equation (4.9) is applied to find that $g^*(u,s) = r(i)h_i^*(s+r(i)u)$. This reduces Equation (4.10) to

$$\begin{aligned}\widetilde{F}_i^*(u,s) &= \sum_j \frac{p_{i,j}}{r(i)} \widetilde{F}_j^*(u,s) r(i) h_i^*(s+r(i)u) + \frac{r(i)}{s+r(i)u} - r(i)H_i^*(s+r(i)u) \\ &= \sum_j p_{i,j} \widetilde{F}_j^*(u,s) h_i^*(s+r(i)u) + \frac{r(i)}{s+r(i)u} - r(i)H_i^*(s+r(i)u). \end{aligned} \tag{4.11}$$

Collecting the $\widetilde{F}_i^*(u,s)$ and $\widetilde{F}_j^*(u,s)$ terms,

$$\widetilde{F}_i^*(u,s)(1 - p_{i,i}h_i^*(s+r(i)u)) - \sum_{j \neq i} p_{i,j} \widetilde{F}_j^*(u,s) h_i^*(s+r(i)u),$$

and applying the matrix operation, $A \circ B$, known as the Hadamard product, where $A \circ B \equiv [a_{i,j}b_{i,j}]$, for all i,j, results in the matrix form,

$$[\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s)]\widetilde{\mathbf{F}}^*(u,s), \tag{4.12}$$

where $\mathbf{I}$ is a K x K identity matrix and

$$\mathbf{h}^*(u,s) = \begin{bmatrix} h_1^*(s+r(1)u) & h_1^*(s+r(1)u) & \cdots & h_1^*(s+r(1)u) \\ h_2^*(s+r(2)u) & h_2^*(s+r(2)u) & \cdots & h_2^*(s+r(2)u) \\ \vdots & \vdots & \ddots & \vdots \\ h_K^*(s+r(K)u) & h_K^*(s+r(K)u) & \cdots & h_K^*(s+r(K)u) \end{bmatrix},$$

$$\mathbf{P} \circ \mathbf{h}^*(u,s) = \begin{bmatrix} p_{1,1}h_1^*(s+r(1)u) & p_{1,2}h_1^*(s+r(1)u) & \cdots & p_{1,K}h_1^*(s+r(1)u) \\ p_{2,1}h_2^*(s+r(2)u) & p_{2,2}h_2^*(s+r(2)u) & \cdots & p_{2,K}h_2^*(s+r(2)u) \\ \vdots & \vdots & \ddots & \vdots \\ p_{K,1}h_K^*(s+r(K)u) & p_{K,2}h_K^*(s+r(K)u) & \cdots & p_{K,K}h_K^*(s+r(K)u) \end{bmatrix}$$

and

$$\widetilde{\mathbf{F}}^*(u,s) = \begin{bmatrix} \widetilde{F}_1^*(u,s) \\ \widetilde{F}_2^*(u,s) \\ \vdots \\ \widetilde{F}_K^*(u,s) \end{bmatrix}.$$

Collecting the remaining terms of Equation (4.11),

$$\frac{r(i)}{(s+r(i)u)} - r(i)H_i^*(s+r(i)u),$$

results in the matrix form

$$[\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r} \tag{4.13}$$

where $\mathbf{A}(u,s) = diag[1/(s+r(1)u), 1/(s+r(2)u), \ldots, 1/(s+r(K)u)]$, $\mathbf{H}^*(u,s) = diag[H_1^*(s+r(1)u), H_2^*(s+r(2)u), \ldots, H_K^*(s+r(K)u)]$, and

$$\mathbf{r} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(K) \end{bmatrix}.$$

Thus, Equation (4.11) in matrix form is

$$[\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s)]\widetilde{\mathbf{F}}^*(u,s) = [\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r}$$
$$\widetilde{F}^*(u,s) = \boldsymbol{\alpha}[\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s)]^{-1}[\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r}. \quad (4.14)$$

If $Z(w)$ depends only on the previous state, then $\psi_{i,j}(w) = p_{i,j}h_i(w)$ and the appropriate Laplace transform of $\psi_{i,j}(w)$, $\psi_{i,j}^*(s+r(i)u) = p_{i,j}h_i^*(s+r(i)u)$, allow us to remove the Hadamard product and rewrite Equation (4.14) as

$$\widetilde{F}^*(u,s) = \boldsymbol{\alpha}[\mathbf{I} - \boldsymbol{\psi}^*(u,s)]^{-1}[\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r} \quad (4.15)$$

where

$$\boldsymbol{\psi}^*(u,s) = \begin{bmatrix} \psi_{1,1}^*(s+r(1)u) & \psi_{1,2}^*(s+r(1)u) & \cdots & \psi_{1,K}^*(s+r(1)u) \\ \psi_{2,1}^*(s+r(2)u) & \psi_{2,2}^*(s+r(2)u) & \cdots & \psi_{2,K}^*(s+r(2)u) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{K,1}^*(s+r(K)u) & \psi_{K,2}^*(s+r(K)u) & \cdots & \psi_{K,K}^*(s+r(K)u) \end{bmatrix}.$$

It is noted that both Equations (4.14) and (4.15) are extremely difficult to evaluate. To demonstrate this difficulty, we examine three specific examples.

4.3 Illustrative Examples

This section provides three specific examples of the lifetime distribution of a single-unit system subject to a homogeneous semi-Markov environment. The first example assumes that environment state holding times are distributed according to an exponential distribution, the second a hyper-exponential distribution and the third an Erlang distribution.

4.3.1 Exponential State Holding Times

The distribution function for an exponential random variable with rate parameter q_i is given by

$$H_i(w) = 1 - e^{-q_i w}$$

and the Laplace transform of $H_i(w)$ evaluated at $s + r(i)u$ is

$$H_i^*(s + r(i)u) = \frac{1}{s + r(i)u} - \frac{1}{s + r(i)u + q_i} \tag{4.16}$$

The probability density function for an exponential(q_i) random variable is

$$h_i(w) = q_i e^{-q_i w}$$

so that

$$h_i^*(s + r(i)u) = \frac{q_i}{(s + r(i)u + q_i)}. \tag{4.17}$$

Additionally, for this example, $p_{i,i} = 0$ and $p_{i,j} = q_{i,j}/q_i$ for $i \neq j$, where $q_i = \sum_{i \neq j} q_{i,j}$. Substituting $p_{i,j}$ and Equations (4.16) and (4.17) into Equation (4.14),

$$\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s) = \begin{bmatrix} 1 & \frac{-q_{1,2}}{s+r(1)u+q_1} & \cdots & \frac{-q_{1,K}}{s+r(1)u+q_1} \\ \frac{-q_{2,1}}{s+r(2)u+q_2} & 1 & \ddots & \frac{-q_{2,K}}{s+r(2)u+q_2} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{-q_{K,1}}{s+r(K)u+q_K} & \frac{-q_{K,2}}{s+r(K)u+q_K} & \cdots & 1 \end{bmatrix} \tag{4.18}$$

and

$$\mathbf{A}(u,s) - \mathbf{H}^*(u,s) = \begin{bmatrix} \frac{1}{s+r(1)u+q_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{s+r(2)u+q_2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{s+r(K)u+q_K} \end{bmatrix}. \tag{4.19}$$

If we rearrange

$$\widetilde{\mathbf{F}}^*(u,s) \;=\; [\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s)]^{-1}[\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r}$$

to

$$[\mathbf{I} - \mathbf{P} \circ \mathbf{h}^*(u,s)]\widetilde{\mathbf{F}}^*(u,s) \;=\; [\mathbf{A}(u,s) - \mathbf{H}^*(u,s)]\mathbf{r}, \tag{4.20}$$

incorporate Equations (4.18) and (4.19), and then simplify, Equation (4.20) is equivalent to the continuous-time Markov chain result shown in [51] and given as

$$\widetilde{F}_i^*(u,s) = \frac{r(i)}{s + r(i)u + q_i} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{q_{i,j}}{s + r(i)u + q_i} \widetilde{F}_j^*(u,s). \tag{4.21}$$

We next consider holding times that are distributed according to a hyper-exponential distribution.

4.3.2 Hyper-exponential State Holding Times

By definition [50], the distribution function of a hyper-exponential random variable with j-phases and rates, $\lambda_{i,j}$, is

$$H_i(w) = \begin{cases} 1 - \sum_{j=1}^{n_i} \alpha_{i,j} e^{-\lambda_{i,j} w}, & w \geq 0 \\ 0, & w < 0 \end{cases} \tag{4.22}$$

which implies that

$$\begin{aligned} H_i^*(s + r(i)u) &= \int_0^\infty e^{-(s+r(i)u)w} \left[1 - \sum_{j=1}^{n_i} \alpha_{i,j} e^{-\lambda_{i,j} w} \right] dw \\ &= \frac{1}{s + r(i)u} - \sum_{j=1}^{n_i} \int_0^\infty e^{-(s+r(i)u)w} \alpha_{i,j} e^{-\lambda_{i,j} w} dw \\ &= \frac{1}{s + r(i)u} - \sum_{j=1}^{n_i} \frac{\alpha_{i,j}}{s + r(i)u + \lambda_{i,j}} \end{aligned} \tag{4.23}$$

where $\alpha_{i,j}$ is the probability that the process enters phase j in state i of the semi-Markov kernel. Additionally, the probability density function for a hyper-exponential random variable is

$$h_i(w) = \begin{cases} \sum_{j=1}^{n_i} \alpha_{i,j} \lambda_{i,j} e^{-\lambda_{i,j} w}, & w \geq 0 \\ 0, & w < 0 \end{cases} \tag{4.24}$$

so that the Laplace transform evaluated at $s + r(i)u$ is given by

$$\begin{aligned} h_i^*(s + r(i)u) &= \int_0^\infty e^{-(s+r(i)u)w} \sum_{j=1}^{n_i} \alpha_{i,j} \lambda_{i,j} e^{-\lambda_{i,j} w} dw \\ &= \sum_{j=1}^{n_i} \frac{\alpha_{i,j} \lambda_{i,j}}{s + r(i)u + \lambda_{i,j}}. \end{aligned} \tag{4.25}$$

By substituting Equations (4.23) and (4.25) into Equation (4.14), $\widetilde{F}^*(u,s)$ reduces to

$$\boldsymbol{\alpha}\begin{bmatrix} \frac{1}{\sum_{j=1}^{n_1}\frac{\alpha_{1,j}\lambda_{1,j}}{s+r(1)u+\lambda_{1,j}}} - p_{1,1} & -p_{1,2} & \cdots & -p_{1,K} \\ -p_{2,1} & \frac{1}{\sum_{j=1}^{n_2}\frac{\alpha_{2,j}\lambda_{2,j}}{s+r(2)u+\lambda_{2,j}}} - p_{2,2} & \ddots & -p_{2,K} \\ \vdots & \ddots & \ddots & \vdots \\ -p_{K,1} & -p_{K,2} & \cdots & \frac{1}{\sum_{j=1}^{n_K}\frac{\alpha_{K,j}\lambda_{K,j}}{s+r(K)u+\lambda_{K,j}}} - p_{K,K} \end{bmatrix}^{-1}\mathbf{r}.$$

This representation provides the 2-dimensional transform of the lifetime distribution of a single-unit system subject to a semi-Markov environment with holding times that have a hyper-exponential distribution. This result is general enough to allow each state of the semi-Markov process to be distributed with a different hyper-exponential distribution. However, numerical inversion of this matrix and the inversion of the Laplace transform are nontrivial. We next show another example, incorporating one of the simplest nonnegative distributions, is equally challenging with respect to obtaining the lifetime distribution.

4.3.3 Erlang State Holding Times

By definition [50], the distribution function for an Erlang random variable with n_i phases and rate parameter λ_i is given by

$$H_i(w) = \begin{cases} 1 - \sum_{j=0}^{n_i-1} e^{-\lambda_i w}\frac{(\lambda_i w)^j}{j!}, & w \geq 0 \\ 0, & w < 0 \end{cases}. \tag{4.26}$$

Making use of the identity

$$\mathcal{L}\left(\frac{w^{n-1}e^{aw}}{(n-1)!}\right) = \frac{1}{(s-a)^n}, \tag{4.27}$$

the Laplace transform of $H_i(w)$ evaluated at $s + r(i)u$ is

$$\begin{aligned} H_i^*(s + r(i)u) &= \frac{1}{s + r(i)u} - \frac{1}{s + r(i)u + \lambda_i} - \frac{\lambda_i}{(s + r(i)u + \lambda_i)^2} - \dots \\ &= \frac{(\lambda_i)^{n_i}}{(s + r(i)u)(s + r(i)u + \lambda_i)^{n_i}}. \end{aligned} \tag{4.28}$$

Additionally, the probability density function for an Erlang(n_i, λ_i) random variable is

$$h_i(w) = \begin{cases} \lambda_i e^{-\lambda_i w} \frac{(\lambda_i w)^{n_i - 1}}{(n_i - 1)!}, & w \geq 0 \\ 0, & w < 0 \end{cases} \tag{4.29}$$

so that

$$h_i^*(s + r(i)u) = \frac{(\lambda_i)^{n_i}}{(s + r(i)u + \lambda_i)^{n_i}}. \tag{4.30}$$

By substituting Equations (4.28) and (4.30) into Equation (4.14), $\widetilde{F}^*(u, s)$ reduces to

$$\boldsymbol{\alpha} \begin{bmatrix} s + r(1)u + \frac{(1-p_{1,1})\lambda_1^{n_1}}{a_{1,n_1}} & \frac{(-p_{1,2})\lambda_1^{n_1}}{a_{1,n_1}} & \dots & \frac{(-p_{1,K})\lambda_1^{n_1}}{a_{1,n_1}} \\ \frac{(-p_{2,1})\lambda_2^{n_2}}{a_{2,n_2}} & s + r(2)u + \frac{(1-p_{2,2})\lambda_2^{n_2}}{a_{2,n_2}} & \ddots & \frac{(-p_{2,K})\lambda_2^{n_2}}{a_{2,n_2}} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{(-p_{K,1})\lambda_K^{n_K}}{a_{K,n_K}} & \frac{(-p_{K,2})\lambda_K^{n_K}}{a_{K,n_K}} & \dots & s + r(K)u + \frac{(1-p_{K,K})\lambda_K^{n_K}}{a_{K,n_K}} \end{bmatrix}^{-1} \mathbf{r}$$

where a_{i,n_i} is obtained recursively by

$$\begin{aligned}
a_{i,1} &= 1 \\
a_{i,2} &= a_{i,1}(s + r(i)u + \lambda_i) + \lambda_i \\
a_{i,3} &= a_{i,2}(s + r(i)u + \lambda_i) + \lambda_i^2 \\
a_{i,4} &= a_{i,3}(s + r(i)u + \lambda_i) + \lambda_i^3 \\
\vdots &\quad \vdots \\
a_{i,n_i} &= a_{i,n_{i-1}}(s + r(i)u + \lambda_i) + \lambda_i^{n_i-1}
\end{aligned} \tag{4.31}$$

It should be noted that if $n_i = 1$, for all i, then $\widetilde{\mathbf{F}}^*(u, s)$ for an Erlang(n_i, λ_i) distribution reduces to the exponential case shown earlier.

The illustrative examples in this section have demonstrated the difficulty in obtaining the lifetime and remaining lifetime distributions. Even if real environment data or degradation data could be incorporated into these analytical results, taking the two-dimensional inverse Laplace transform to obtain the lifetime distribution would be nontrivial. These results allow greater flexibility in modelling real world systems than the Markovian environment examined in Chapter III. The results of Chapter V provide a way to circumvent the analytical difficulties of Chapter IV and allow us to utilize the analytical results in Chapter III. In particular, it will be shown that a Markovian environment can be constructed via phase-type distribution approximations in order to facilitate implementation of Equations (3.51) and (3.67).

V. Phase-Type Approximations

Numerical implementation of the semi-Markov environment results of Chapter IV is nontrivial, even if the holding time distributions are known. To remedy this problem, we turn our attention to phase-type (PH) distributions which can approximate the holding time distributions in the homogeneous semi-Markovian environment. Moreover, PH-distributions retain the Markov property which allows us to utilize the analytical lifetime distribution result of Equation (3.51) while the system still operates in a semi-Markovian environment. Another key point is that PH-distribution approximations are easily determined from both observable environment and observable degradation data. Hence, these approximations may be used to bypass the analytical results contained in Chapter IV.

This chapter begins by introducing PH-distributions and their main properties in addition to some specific types of PH-distributions. In Section 5.2, we explore PH-distribution approximations to general distributions by highlighting a few alternative approaches and comparing these procedures numerically. Our purpose is to choose the "best" PH-distribution approximation in the sense of a minimal representation. Following this numerical examination, we estimate the lifetime distribution utilizing PH-distribution approximations of state holding time distributions using only real, observed data.

5.1 Properties of PH-Distributions

The class of matrix-exponential (ME) distributions discussed in Section 3.3.1 contains all PH-distributions. Hence, PH-distributions also possess a rational Laplace transform, but Neuts [65] indicates PH-distributions do not contain complex parameters. The majority of the properties contained in this section can be found in [65], but are restated here for completeness.

Define a homogeneous continuous-time Markov process having $m+1$ states with infinitesimal generator

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \mathbf{T} & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{bmatrix}, \tag{5.1}$$

and $(\boldsymbol{\alpha}, \alpha_{m+1})$ as the initial probability distribution of $\boldsymbol{\Upsilon}$ where $\boldsymbol{\alpha}$ is a row vector. For the $m \times m$ matrix $\mathbf{T}$, $T_{i,i} < 0$ and $T_{i,j} \geq 0, i \neq j$. Additionally, $\mathbf{T}\boldsymbol{e} + \mathbf{T}^0 = \mathbf{0}$ and $\boldsymbol{\alpha}\boldsymbol{e} + \alpha_{m+1} = 1$ where $\mathbf{T}^0$ is a column vector and $\boldsymbol{e}$ is a column vector of ones. In this formulation, Neuts [65] assumes the first m states are transient and that the final state, $m+1$, is absorbing where absorption is certain to occur in a finite amount of time.

Definition 4. *The probability distribution, $F(\cdot)$, of the time until absorption into state $m+1$, with associated initial probability distribution, $(\boldsymbol{\alpha}, \alpha_{m+1})$ of $\boldsymbol{\Upsilon}$ is*

$$F(w) = 1 - \boldsymbol{\alpha} \exp(\mathbf{T}w)\boldsymbol{e}. \tag{5.2}$$

Definition 5. *The probability distribution, $F(w)$, on $[0, \infty)$ is a phase-type distribution if and only if it is the time until absorption in a finite Markov process as that shown in Equation (5.1). This PH-distribution has representation $PH(\boldsymbol{\alpha}, \mathbf{T})$.*

Neuts [65] indicates that the noncentral moments of a PH-distribution are the same as those for the the ME-distribution in Equation (3.56). We will make use of one theorem that Neuts [65] provides on PH-distributions.

Theorem 8. *A finite mixture of PH-distributions is a PH-distribution.*

We provide examples of a general PH-distribution and three special cases of the general PH-distribution with both a visual [70] and mathematical representation to help clarify the differences. Additional PH-distributions referenced in subsequent

sections provide only the mathematical representation. A three-phase, general phase-type distribution is shown in Figure 5.1 where $p_{i,j}$ is the probability that the finite Markov process transitions from state i to state j. Additionally, upon entering state i, an exponential amount of time, with rate λ_i, is spent in that state before transitioning to the next state. The process ends upon entering the absorbing state. Figure 5.1 displays a PH-distribution with representation $PH(\boldsymbol{\alpha}, \mathbf{T})$ where $\boldsymbol{\alpha} = [p_{0,1} \quad p_{0,2} \quad p_{0,3}]$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & p_{1,2}\lambda_1 & p_{1,3}\lambda_1 \\ p_{2,1}\lambda_2 & -\lambda_2 & p_{2,3}\lambda_2 \\ p_{3,1}\lambda_3 & p_{3,2}\lambda_3 & -\lambda_3 \end{bmatrix}. \tag{5.3}$$

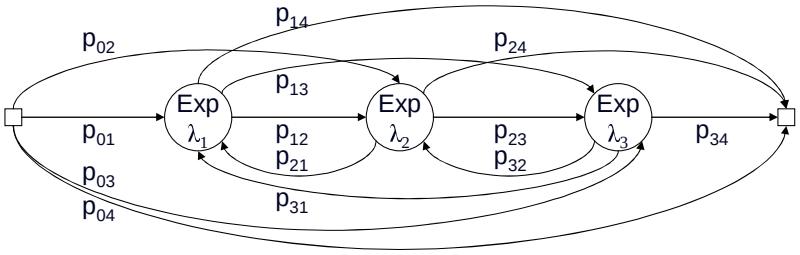

Figure 5.1: A three-phase general PH-distribution.

A three-phase acyclic PH-distribution is graphically shown in Figure 5.2 where $\boldsymbol{\alpha} = [p_{0,1} \quad p_{0,2} \quad p_{0,3}]$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & p_{1,2}\lambda_1 & p_{1,3}\lambda_1 \\ 0 & -\lambda_2 & p_{2,3}\lambda_2 \\ 0 & 0 & -\lambda_3 \end{bmatrix}. \tag{5.4}$$

Acyclic PH-distributions are similar to general PH-distributions in that $p_{i,j}$ is the probability of transitioning from state i to state j and an exponential amount of

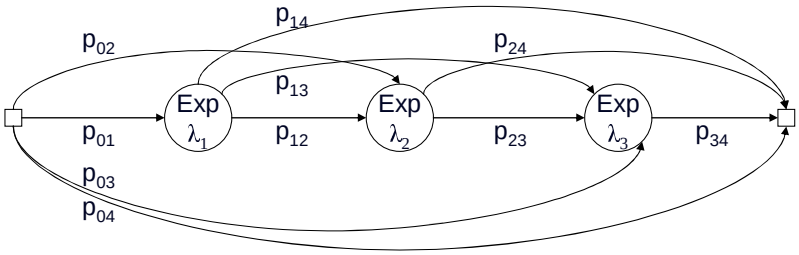

Figure 5.2: A three-phase acyclic PH-distribution.

time, with rate λ_i, is spent in state i before transitioning to the next state. However, this PH-distribution [70] does not allow transitions from state i to state j if $i > j$. A special case of an acyclic PH-distribution is the Coxian distribution. This case allows transitions only to the adjacent, higher state or the absorbing state. A three-phase Coxian distribution is shown in Figure 5.3 where $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & p_{1,2}\lambda_1 & 0 \\ 0 & -\lambda_2 & p_{2,3}\lambda_2 \\ 0 & 0 & -\lambda_3 \end{bmatrix}. \tag{5.5}$$

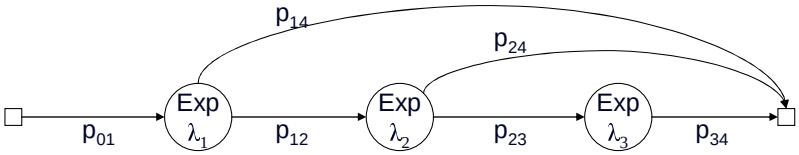

Figure 5.3: A three-phase Coxian distribution.

Lastly, another type of acyclic PH-distribution, and a subset of the Coxian distributions, is the k-phase Erlang distribution where the amount of time spent in each state is exponentially distributed with rate λ and only the first and last state

allow transitions to the absorbing state. A three-phase example is shown in Figure 5.4 where $\boldsymbol{\alpha} = [1 \quad 0 \quad 0]$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda & p_{1,2}\lambda & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{bmatrix}. \tag{5.6}$$

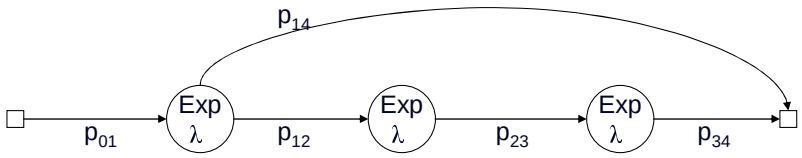

Figure 5.4: A three-phase Erlang distribution.

The acyclic, Coxian, and k-phase Erlang distributions are discussed in greater detail in the Section 5.2.

5.2 Approximating a General Distribution

This section provides the necessary background to choose an appropriate PH-distribution approximation method required for the transformation of the semi-Markovian environment into a Markovian environment in order to utilize the analytical lifetime distribution. We examine the PH-distribution approximation methods, minimal PH-distribution representations, reasonings for choosing a moment-matching method, and then compare the various moment-matching methods. This section demonstrates that if we have holding time observations, which are obtainable from an observable environment and observable degradation, then we can obtain a PH-distribution approximation. Section 5.3 establishes that the individual PH-

distribution approximations explored in this section are useful for transforming the homogeneous semi-Markovian environment into a homogeneous Markovian environment where the analytical lifetime distribution result can be estimated, resulting in extended applicability. This section first summarizes various methods to obtain a PH-distribution approximation.

5.2.1 Summary of Methods

Choosing an appropriate PH-distribution approximation is not an easy task and one must carefully examine the literature for techniques and comparisons. Osogami and Harchol-Balter [70], Perros [73], Johnson [36], and Altiok [5] provide excellent summaries of techniques to obtain PH-distribution approximations. These techniques rely mostly upon an estimate of the squared coefficient of variation (c^2) of the distribution, defined as the variance divided by the mean squared. The range for c^2 is broken into three intevals, $0 < c^2 \leq 0.5$, $0.5 < c^2 \leq 1.0$, and $c^2 > 1.0$. Once the c^2 value is estimated, various techniques, including the method of moments, maximum likelihood, and minimum distance, are used to estimate the PH-distribution parameters. The method of moments is further subdivided into mainly either two- or three-moment methods while nonlinear programming techniques are often associated with maximum likelihood and minimum distance techniques. There is little dispute that a k-phase Erlang [4] distribution provides the "best" approximation for $c^2 < 0.5$ where best is normally synonymous with minimal representation. However, the best method and type of PH-distribution approximation for $c^2 > 0.5$ is not as obvious.

In order to use the minimum distance estimation, Perros [73] indicated that sample observations from the distribution to be approximated are required to find this PH-distribution approximation. However, for our purposes, the unit lifetime distribution is unknown; hence, we cannot employ this technique. Additionally, Lang and Arthur [53] provided a comparison of PH-distribution approximations us-

ing two moment-matching and two maximum-likelihood parameter estimation techniques. The overall comparison favors the moment-matching techniques. Perros [73] indicated that the maximum likelihood and minimum distance techniques require nonlinear optimization. Johnson [36] pointed out that many of the problems associated with maximum likelihood techniques, such as multiple global optima and slow convergence, have been encountered in fitting PH-distributions. For these reasons, we select moment matching as our technique to estimate the parameters of a PH-distribution approximation.

With this moment matching technique comes some confusion as to the number of moments that should be matched as well as the type of PH-distribution approximation. It is generally better to match as many moments as possible since two distributions are equivalent if all moments match. However, as the number of moments increases, the data requirements increase as do the number of phases required for the approximation. Osogami and Harchol-Balter's [70] provide the best overall summary of these techniques although they indicated Marie [58] used two moments to match a generalized Erlang for $c^2 < 1$ and also used two moments to match a two-phase Coxian for $c^2 \geq 0.5$. Perros [73] also indicated that Marie [58] advocated a two-phase Coxian for $c^2 \geq 0.5$. Marie [58] provided a two moment matching algorithm for a k-phase Erlang approximation if $0 < c^2 \leq 0.5$ and a two-phase Coxian approximation if $0.5 < c^2 < 1.0$. Altiok [5] and Telek and Heindl [86] provided three moment matching algorithms if $c^2 > 1$ where Altiok used a two-phase Coxian approximation and Telek and Heindl [86] used a two-phase canonical acyclic PH-distribution. Initially, the two-phase canonical acyclic PH-distribution transitions to state 1 with probability p or transitions to state 2 with probability $1 - p$. If the transition to state 1 occurs, then an exponential amount of time, with rate λ_1 is spent in state one and then an exponential amount of time, with rate λ_2 is spent in state two before absorption. If the transition to state two occurs initially, then an exponential amount of time, with rate λ_2 is spent in state two before absorption.

This two-phase canonical acyclic PH-distribution [86] has $\boldsymbol{\alpha} = \begin{bmatrix} p & 1-p \end{bmatrix}$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$$

where $0 \leq p \leq 1$ and $0 < \lambda_1 \leq \lambda_2$. Osogami and Harchol-Balter [70] provided a three-moment matching algorithm which covered most all nonnegative distributions using an Erlang-Coxian PH-distribution approximation. This N-phase Erlang-Coxian PH-distribution [70] combined a k-phase Erlang and a two-phase Coxian with $\boldsymbol{\alpha} = \begin{bmatrix} p_1 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}$ and

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\lambda_1 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_2 & p_2\lambda_2 \\ 0 & 0 & 0 & 0 & 0 & -\lambda_3 \end{bmatrix}$$

where $0 < p_1 \leq 1$ and $0 \leq p_2 \leq 1$. It should be noted that if $p_1 \neq 1$, then there will be an initial jump in probability at 0. Before we compare these various PH-distribution approximations, we address the concept of a minimal PH-distribution as both Osogami and Harchol-Balter [70] and Telek and Heindl [86] address it in their articles.

5.2.2 Minimal PH-Distribution

Our intention is to provide a technique that can be implemented quickly to obtain the lifetime and remaining lifetime distributions for systems prognosis. Hence, a minimal PH-distribution is desired because the smallest representation requires

the fewest estimated parameters. However, the literature can be confusing on what constitutes a minimal representation.

Neuts [65] defines a minimal PH-distribution as one whose matrix $\mathbf{T}$ is as small as possible and indicated that the minimal representation for a PH-distribution is a difficult, unsolved problem. However, Osogami and Harchol-Balter [70] recently mapped general distributions to minimal PH-distributions. Both are correct in their statements, however. Osogami and Harchol-Balter [70] clarify in their article that their resulting distribution is minimal with respect to a specific acyclic PH-distribution when matching three moments, and hence, is not minimal with respect to the general PH-distribution to which Neuts [65] referred. Additionally, Telek and Heindl [86] examined canonical forms [21] which are a subclass of acyclic PH-distributions that also admit minimal representations. Hence, neither of these recent techniques consider general PH-distributions; they both restrict their method to a subset of PH-distribution (acyclic PH-distributions) and then restrict their method even further by considering only the first three moments. Those who claim a minimal PH-distribution representation most often use some form of an acyclic PH-distribution and not a general PH-distribution.

Since it is important to use a minimal or near-minimal PH-distribution in our technique, we continued to examine minimal acyclic PH-distributions. Altiok [5] does not address minimal representation in his article, but since Marie's [58] two-moment approximation for $0 < c^2 \leq 0.5$ is in line with [4], it provided the minimal representation with respect to two moments. Telek and Heindl's [86] three-moment approximation also provides a minimal representation and Osogami and Harchol-Balter's [70] three-moment approximation provides near minimal representations. We next examine procedures to choose between these techniques.

5.2.3 Choosing an Appropriate PH-Distribution

If a general distribution has $c^2 < 0.5$, a k-phase Erlang distribution [58] should be used via the first two non-central moments, μ_1 and μ_2, where

$$c^2 = \frac{\mu_2 - \mu_1^2}{\mu_1^2}, \tag{5.7}$$

and

$$\frac{1}{k} \leq c^2 < \frac{1}{k-1}. \tag{5.8}$$

Aldous and Shepp [4] showed that this value, k, is the minimal number of phases in an acyclic PH-distribution. Marie [58] used c^2 and k to find the two remaining parameters, a and λ_1, of the associated k-phase Erlang distribution where

$$a = 1 - \frac{2kc^2 + k - 2 - (k^2 + 4 - 4kc^2)^{1/2}}{2(c^2+1)(k-1)} \tag{5.9}$$

and

$$\lambda_1 = \frac{1 + (k-1)a}{\mu_1} \tag{5.10}$$

resulting in the $k \times k$ matrix

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & a\lambda_1 & 0 & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & 0 & -\lambda_1 & \lambda_1 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & -\lambda_1 \end{bmatrix}. \tag{5.11}$$

For clarification, Marie [58] referred to the representation of Equation (5.11) as a generalized k-phase Erlang whereas Neuts [65] uses the representation

$$\mathbf{T}_{Neuts} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & -\lambda_k \end{bmatrix}. \tag{5.12}$$

The difference between these two representations is quite significant as Neuts' [65] representation cannot transition to the absorbing state until the final phase whereas Marie's [58] representation can transition to the absorbing state from the initial and final phase. Additionally, the amount of time spent in each state is exponentially distributed with the same rate parameter, λ_1, in Marie's [58] representation, but Neuts' [65] representation has a different rate for each state. This dissertation will use Marie's representation (Equation (5.11)) of the generalized k-phase Erlang distribution when $c^2 < 0.5$.

If a general distribution has $0.5 \leq c^2 \leq 1$, Telek and Heindl's [86] approach and Marie's [58] approach provide a PH-distribution approximation. Marie [58] matched two moments of the general distribution with a two-phase Coxian distribution and Telek and Heindl [86] matched three moments of the general distribution to their two-phase canonical acyclic PH-distribution. However, for their method to work in this range of c^2, Telek and Heindl [86] require bounds on the third moment of the form

$$3\mu_1^3(3c^2 - 1 + \sqrt{2}(1-c^2)^{\frac{3}{2}}) \leq \mu_3 \leq 6\mu_1^3 c^2. \tag{5.13}$$

If the third moment does not fall within these bounds, the method should not be used. There were no bounds on Marie's approach to estimate the three parameters

using the first two noncentral moments, μ_1 and μ_2, where

$$\lambda_1 = \frac{2}{\mu_1}, \quad \lambda_2 = \frac{1}{\mu_1 c^2}, \quad a = \frac{1}{2c^2} \tag{5.14}$$

resulting in a two-phase Coxian distribution where

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & a\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}. \tag{5.15}$$

Lastly, if a general distribution has $c^2 > 1$, Sauer and Chandy [79] used a two-branch hyper-exponential distribution after matching two moments. Whitt [93] also used a two-branch hyper-exponential distribution, Altiok [5] a two-phase Coxian distribution, and Telek and Heindl [86] a two-phase canonical acyclic PH-distribution after matching three moments. Whitt [93] showed that for $c^2 > 1$, it is best to match at least three moments to reduce the maximum relative error between the true and approximated distributions. Consequently, we do not recommend using the method of Sauer and Chandy [79] which matches only two moments. Additionally, Whitt's [93] hyper-exponential distribution requires three parameter estimates as well as the determination of the parameter, r, which is estimated from the first three parameter estimates. Since Telek and Heindl's method and Altiok's method required only three parameter estimates, we decided to examine their methods to approximate a general distribution with $c^2 > 1$. However, both methods place the same restriction on the true distribution's third moment, namely

$$\mu_3 > \frac{3(c^2+1)^2\mu_1^3}{2}. \tag{5.16}$$

If this restriction is not satisfied, neither method is appropriate.

Osogami and Harchol-Balter's [70] moment matching techniques are not based on the squared coefficient of variation, but do require estimates for the first three

moments. They fit three moments for $c^2 > 0$ and provide three methods, a simple closed-form method, an improved closed-form method, and a numerically stable closed-form method. Let the minimal number of phases for their acyclic PH-distribution that matches the first three moments be ϑ. Their simple close-form method has at most, $\vartheta + 2$ phases, their improved closed-form method has at most, $\vartheta + 1$ phases, and their numerically stable closed-form method has at most, $\vartheta + 2$ phases. Their Erlang-Coxian (EC) distribution is able to match almost all nonnegative distributions. This claim is based upon the second and third normalized moments, m_2 and m_3, where

$$\begin{aligned} m_2 &= \frac{u_2}{u_1^2} \\ m_3 &= \frac{u_3}{u_1 u_2}. \end{aligned} \tag{5.17}$$

It was shown [37] that a PH-distribution and a general distribution agree on their first three moments if and only if $m_3 > m_2 > 1$. Additionally, all nonnegative distributions satisfy $m_3 \geq m_2 \geq 1$ which justifies this claim (see [40]).

We chose to numerically compare those moment-matching techniques highlighted in this subsection that provide a minimal, or near minimal, representation of an acyclic PH-distribution. Specifically, we compare the PH-distribution approximation techniques of Marie [58], Altiok [5], Telek and Heindl [86], and Osogami and Harchol-Balter [70].

5.2.4 Comparison of Techniques

We compare PH-distribution approximations of Weibull, beta, and gamma distributions over the ranges $0 \leq c^2 \leq 0.5$, $0.5 < c^2 < 1$, and $c^2 > 1$. For the range, $0 \leq c^2 \leq 0.5$, we compare Marie's [58] and Osogami and Harchol-Balter's [70] PH-distribution approximations with known Weibull, beta, and gamma distributions. For the range, $0.5 < c^2 < 1$, we compare Marie's [58], Telek and Heindl's [86],

and Osogami and Harchol-Balter's [70] PH-distribution approximations with known Weibull, beta, and gamma distributions. Finally, for $c^2 > 1$, we compare Altiok's [5], Telek and Heindl's [86], and Osogami and Harchol-Balter's [70] PH-distribution with known Weibull, beta, and gamma distributions. Table 5.1 provides the general distributions, their noncentral moments, and their squared coefficients of variation. All moments were computed using the Matlab® computing environment. In Table 5.1, Weibull(a, b) represents a Weibull distribution with scale parameter a and shape parameter b, Beta(a, b) represents a beta distribution with shape parameters a and b, and Gamma(a, b) represents a gamma distribution with shape parameter a and scale parameter b. We use these distributions to form our comparisons.

Table 5.1: Lower moments and c^2 for various distributions.

Distribution	μ_1	μ_2	μ_3	c^2
Weibull(2.5, 3.0)	0.658	0.490	0.400	0.132
Beta(2.0, 4.0)	0.333	0.143	0.071	0.286
Gamma(2.5, 2.0)	5.000	35.000	315.000	0.400
Beta(1.2, 4.0)	0.231	0.082	0.036	0.538
Gamma(1.3, 2.0)	2.600	11.960	78.936	0.769
Weibull(3.0, 1.03)	0.340	0.225	0.219	0.943
Gamma(0.7, 2.0)	1.400	4.760	25.704	1.429
Weibull(3.0, 0.6)	0.241	0.238	0.494	3.091
Beta(0.1, 2.0)	0.048	0.017	0.009	6.452

Approximation Comparisons ($0 < c^2 < 0.5$)

This section provides detailed comparisons for Weibull(2.5, 3.0), beta(2.0, 4.0), and gamma(2.5, 2.0) distributions shown in Table 5.1. The PH-distribution approximations are those due to Marie [58] and Osogami and Harchol-Balter [70].

For the Weibull(2.5, 3.0) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = [1 \quad 0 \quad \cdots \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -12.07 & 11.98 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -12.07 & 12.07 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -12.07 & 12.07 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -12.07 & 12.07 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -12.07 & 12.07 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -12.07 & 12.07 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -12.07 & 12.07 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -12.07 \end{bmatrix} \tag{5.18}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has a 1×10 row vector $\boldsymbol{\alpha} = [0.977 \quad 0 \quad \ldots \quad 0]$ and a 10×10 $\mathbf{T}$ matrix,

$$\begin{bmatrix} -14.02 & 14.02 & 0 & & \cdots & 0 \\ 0 & -14.02 & 14.02 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -14.02 & 14.02 & 0 \\ 0 & \ldots & & 0 & -25.97 & 21.55 \\ 0 & \ldots & & & 0 & -12.87 \end{bmatrix}. \tag{5.19}$$

Figure 5.5 shows the actual distribution comparisons while Table 5.2 compares the number of phases and the maximum absolute deviation (MAD) in probability.

Table 5.2: Weibull(2.5, 3.0) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.132	8	0.0354
Osogami	0.132	10	0.0399

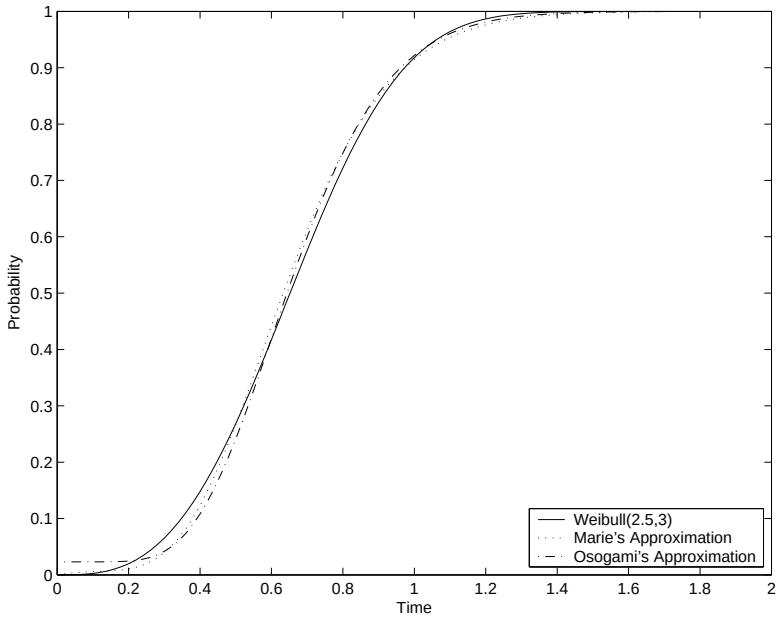

Figure 5.5: Phase-type approximation of a Weibull(2.5, 3) cdf.

For the beta(2.0, 4.0) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = [1 \quad 0 \quad 0 \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -11.58 & 11.05 & 0 & 0 \\ 0 & -11.58 & 11.58 & 0 \\ 0 & 0 & -11.58 & 11.58 \\ 0 & 0 & 0 & -11.58 \end{bmatrix} \tag{5.20}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = [0.933 \quad 0 \quad 0 \quad 0 \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -14 & 14 & 0 & 0 & 0 \\ 0 & -14 & 14 & 0 & 0 \\ 0 & 0 & -14 & 14 & 0 \\ 0 & 0 & 0 & -14 & 14 \\ 0 & 0 & 0 & 0 & -14 \end{bmatrix}. \tag{5.21}$$

Figure 5.6 shows the actual distribution comparisons while Table 5.3 compares the number of phases and the maximum absolute deviation (MAD) in probability.

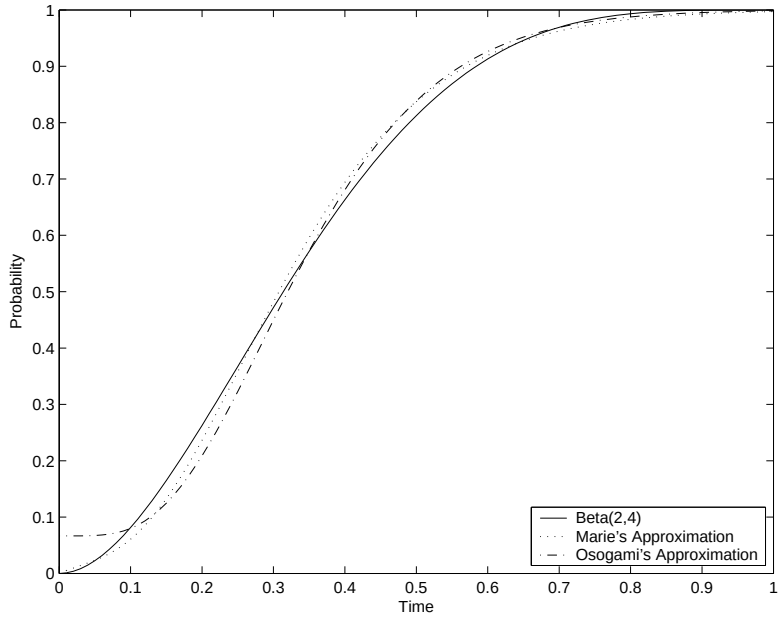

Figure 5.6: Phase-type approximation of a Beta(2.0, 4.0) cdf.

Table 5.3: Beta(2.0, 4.0) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.286	4	0.0324
Osogami	0.286	5	0.0667

For the gamma(2.5, 2.0) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = [1 \quad 0 \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -0.562 & 0.508 & 0 \\ 0 & -0.562 & 0.562 \\ 0 & 0 & -0.562 \end{bmatrix} \tag{5.22}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = [1 \quad 0 \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -0.500 & 0.500 & 0 \\ 0 & -0.877 & 0.744 \\ 0 & 0 & -0.456 \end{bmatrix}. \tag{5.23}$$

Figure 5.7 shows the actual distribution comparisons while Table 5.4 compares the number of phases and the maximum absolute deviation (MAD) in probability.

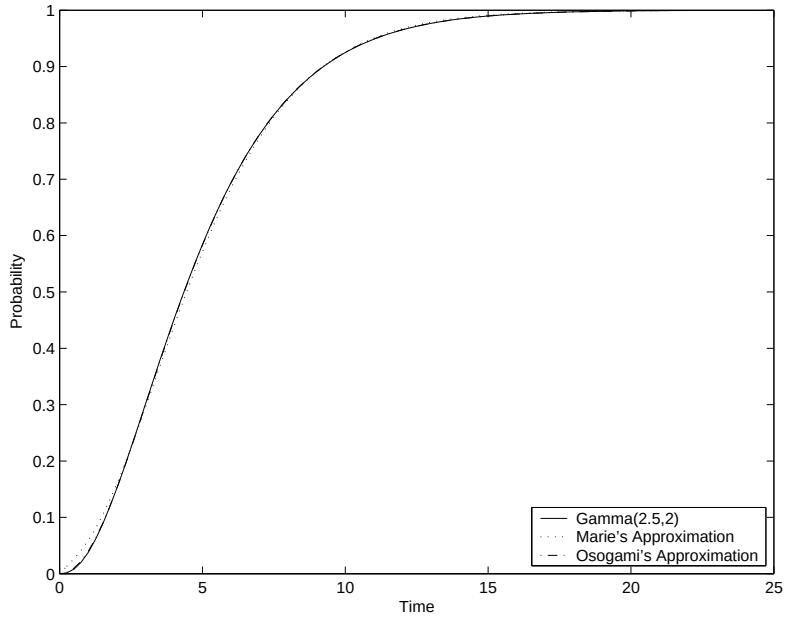

Figure 5.7: Phase-type approximation of a Gamma(2.5, 2.0) cdf.

Table 5.4: Gamma(2.5, 2.0) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.400	3	0.0214
Osogami	0.400	3	0.0020

We next examine the comparisons for $0.5 < c^2 < 1.0$.

Approximation Comparisons ($0.5 < c^2 < 1.0$)

This section provides detailed comparisons for beta(1.2, 4.0), gamma(1.3, 2.0), and Weibull(3.0, 1.03) distributions shown in Table 5.1. The PH-distribution approximations are provided in Marie [58], Telek and Heindl [86], and Osogami and Harchol-Balter [70].

For the beta(1.2, 4.0) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -8.67 & 8.06 \\ 0 & -8.06 \end{bmatrix}. \tag{5.24}$$

Telek and Heindl's [86] PH-distribution approximation is not applicable because their third moment bounds were violated. Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.87 & 0 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -11.16 & 11.16 & 0 \\ 0 & -12.32 & 12.17 \\ 0 & 0 & -10.42 \end{bmatrix}. \tag{5.25}$$

Figure 5.8 shows the actual distribution comparisons while Table 5.5 compares the number of phases and the maximum absolute deviation (MAD) in probability.

Table 5.5: Beta(1.2, 4.0) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.538	2	0.0399
Telek		N/A	N/A
Osogami	0.538	3	0.1299

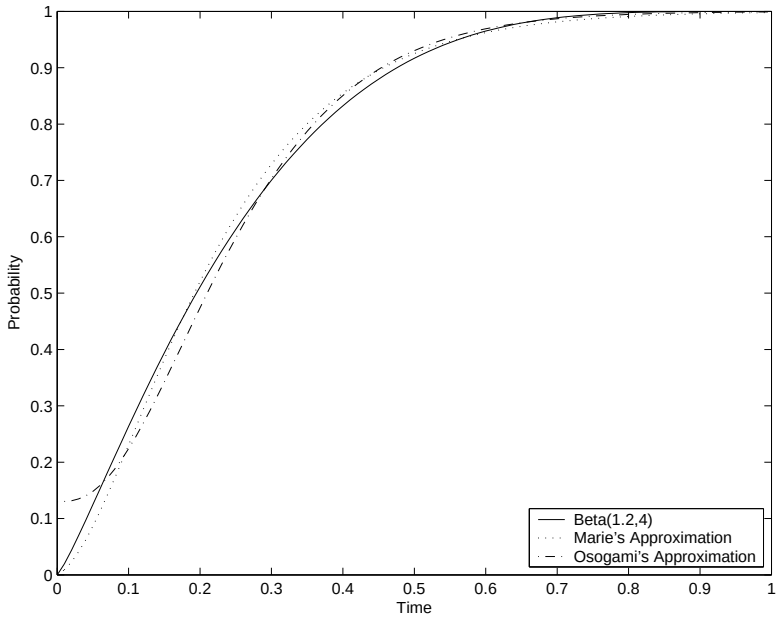

Figure 5.8: Phase-type approximation of a Beta(1.2, 4.0) cdf.

For the gamma(1.3, 2.0) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -0.77 & 0.5 \\ 0 & -0.5 \end{bmatrix} \tag{5.26}$$

and Telek and Heindl's [86] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.78 & 0.22 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -0.47 & 0.47 \\ 0 & -1.07 \end{bmatrix}, \tag{5.27}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -1.07 & 0.84 \\ 0 & -0.47 \end{bmatrix}. \tag{5.28}$$

Figure 5.9 shows the actual distribution comparisons while Table 5.6 compares the number of phases and the maximum absolute deviation (MAD) in probability.

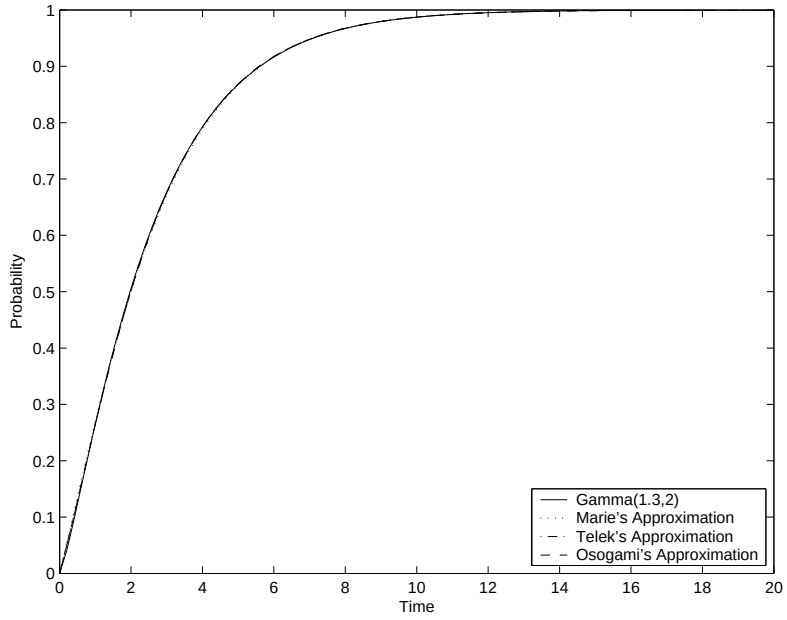

Figure 5.9: Phase-type approximation of a Gamma(1.3, 2.0) cdf.

Table 5.6: Gamma(1.3, 2.0) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.769	2	0.0150
Telek	0.769	2	0.0086
Osogami	0.769	2	0.0086

For the Weibull(3.0, 1.03) distribution, Marie's [58] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -5.88 & 3.12 \\ 0 & -3.12 \end{bmatrix} \tag{5.29}$$

and Telek and Heindl's [86] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.524 & 0.476 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -3.12 & 3.12 \\ 0 & -5.81 \end{bmatrix}, \tag{5.30}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.987 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -8.51 & 5.99 \\ 0 & -3.10 \end{bmatrix}. \tag{5.31}$$

Figure 5.10 shows the actual distribution comparisons while Table 5.7 compares the number of phases and the maximum absolute deviation (MAD) in probability.

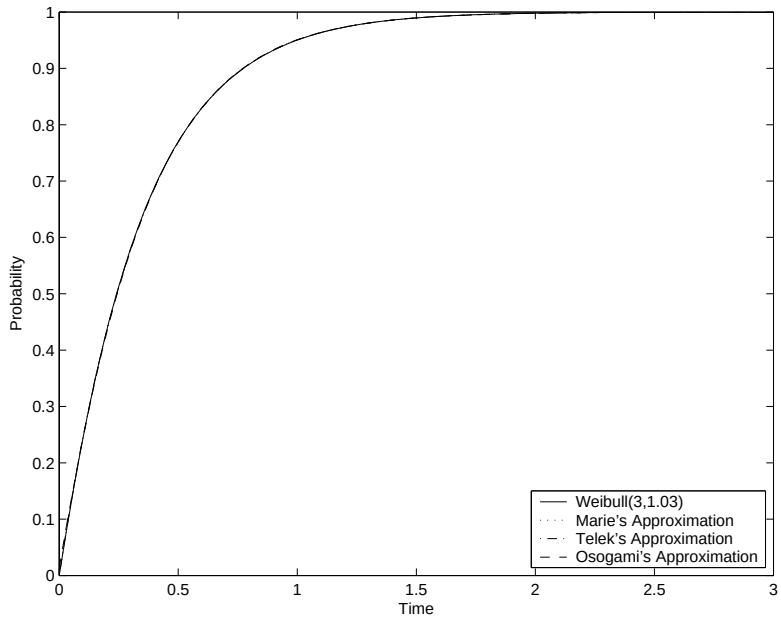

Figure 5.10: Phase-type approximation of a Weibull(3.0, 1.03) cdf.

We next examine the comparisons for $c^2 > 1$.

Table 5.7: Weibull(3.0, 1.03) versus PH approximations.

Method	c^2	Phases	MAD
Marie	0.943	2	0.0022
Telek	0.943	2	0.0022
Osogami	0.943	2	0.0133

Approximation Comparisons ($c^2 > 1.0$)

This section provides detailed comparisons for gamma(0.7, 2.0), Weibull(3.0, 0.6), and beta(0.1, 2.0) distributions shown in Table 5.1. The PH-distribution approximations provided in Altiok [5], Telek and Heindl [86], and Osogami and Harchol-Balter [70].

For the gamma(0.7, 2.0) distribution, Altiok's [5] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -2.31 & 1.22 \\ 0 & -0.55 \end{bmatrix} \tag{5.32}$$

and Telek and Heindl's [86] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.527 & 0.473 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -0.55 & 0.55 \\ 0 & -2.31 \end{bmatrix}, \tag{5.33}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -2.31 & 1.22 \\ 0 & -0.55 \end{bmatrix}. \tag{5.34}$$

Figure 5.11 shows the actual distribution comparisons while Table 5.8 compares the number of phases and the maximum absolute deviation (MAD) in probability.

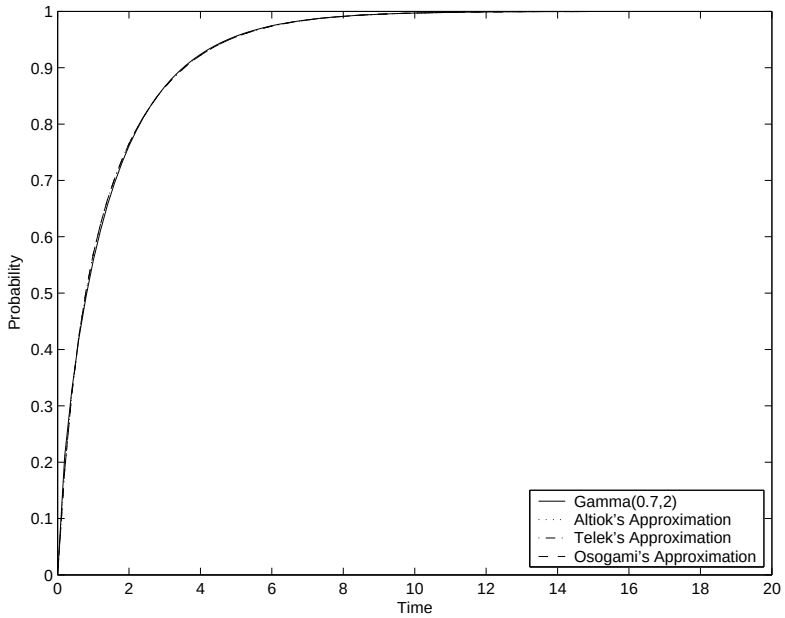

Figure 5.11: Phase-type approximation of a Gamma(0.7, 2.0) cdf.

Table 5.8: Gamma(0.7, 2.0) versus PH approximations.

Method	c^2	Phases	MAD
Altiok	1.429	2	0.0248
Telek	1.429	2	0.0248
Osogami	1.429	2	0.0248

For the Weibull(3.0, 0.6) distribution, Altiok's [5] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -8.08 & 1.25 \\ 0 & -1.32 \end{bmatrix} \tag{5.35}$$

and Telek and Heindl's [86] PH-distribution approximation has $\boldsymbol{\alpha} = \begin{bmatrix} 0.155 & 0.845 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -1.32 & 1.32 \\ 0 & -8.08 \end{bmatrix}, \tag{5.36}$$

whereas Osogami and Harchol-Balter's [70] PH-distribution approximation has $\boldsymbol{\alpha} = [1 \quad 0]$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -8.08 & 1.25 \\ 0 & -1.32 \end{bmatrix}. \tag{5.37}$$

Figure 5.12 shows the actual distribution comparisons while Table 5.9 compares the number of phases and the maximum absolute deviation (MAD) in probability.

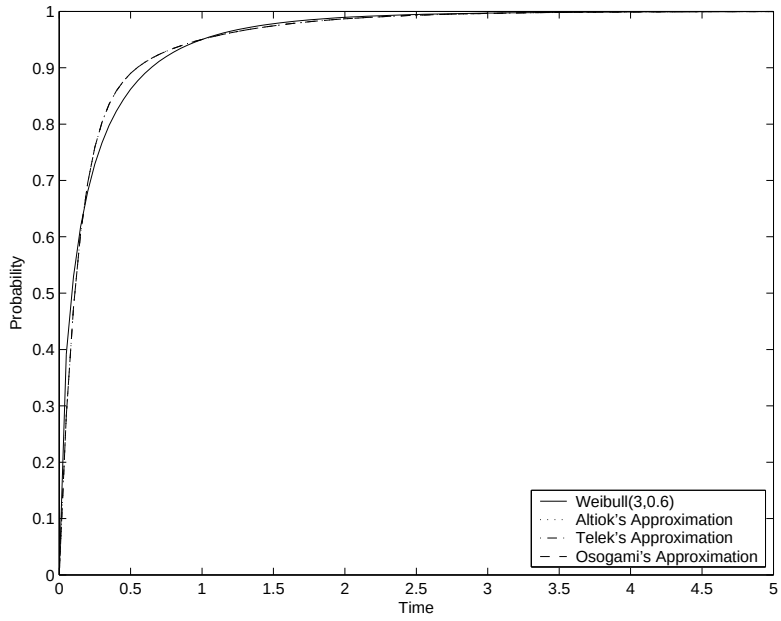

Figure 5.12: Phase-type approximation of a Weibull(3.0, 0.6) cdf.

Table 5.9: Weibull(3.0, 0.6) versus PH approximations.

Method	c^2	Phases	MAD
Altiok	3.091	2	0.1090
Telek	3.091	2	0.1090
Osogami	3.091	2	0.1090

For the Beta(0.1, 2.0) distribution, Altiok's [5] and Telek and Heindl's [86] PH-distribution approximations do not exist because their 3^{rd} moment bounds were violated, whereas Osogami and Harchol-Balter's [70] PH-distribution approximation

has $\boldsymbol{\alpha} = \begin{bmatrix} 0.241 & 0 \end{bmatrix}$ with $\mathbf{T}$ matrix

$$\begin{bmatrix} -14.24 & 10.90 \\ 0 & -6.01 \end{bmatrix}. \tag{5.38}$$

Figure 5.13 shows the actual distribution comparisons while Table 5.10 compares the number of phases and the maximum absolute deviation (MAD) in probability.

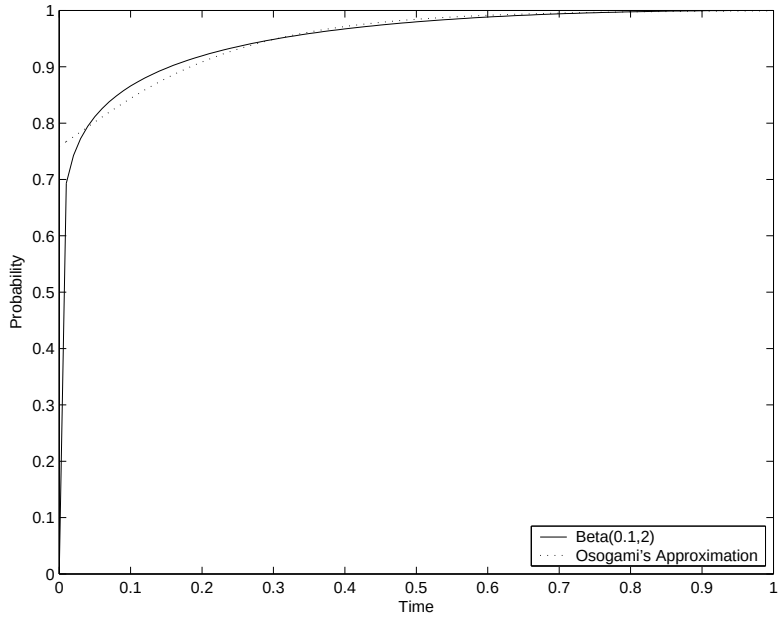

Figure 5.13: Phase-type approximation of a Beta(0.1, 2.0) cdf.

Table 5.10: Beta(0.1, 2.0) versus PH approximations.

Method	c^2	Phases	MAD
Altiok		DNE	DNE
Telek		DNE	DNE
Osogami	6.452	2	0.7589

Table 5.11 contains a summary of all comparisons where it is important to note that Altiok's [5] and Telek and Heindl's [86] approximation techniques did not provide an approximation in a few cases because the third moment restrictions were violated. Additionally, the maximum absolute deviation in probability is not neces-

sarily a good measure because those approximation methods marked by an asterisk, specifically Osogami and Harchol-Balter's [70] PH-distribution approximations, have a probability mass at time 0. Thus, when comparing individual distributions, there does not appear to be a PH-distribution approximation technique that clearly provides a better approximation over the other techniques when comparing individual distributions. However, the method we present in the next section combines PH-distribution approximations to obtain the lifetime distribution

Table 5.11: Summary of PH-distribution comparisons.

Distribution	Approximation Method	c^2	Phases	MAD
	Marie		8	0.0354
Weibull(2.5, 3.0)	Osogami*	0.132	10	0.0399
	Marie		4	0.0324
Beta(2.0, 4.0)	Osogami*	0.286	5	0.0667
	Marie		3	0.0214
Gamma(2.5, 2.0)	Osogami	0.400	3	0.0020
	Marie		2	0.0399
	Telek		N/A	N/A
Beta(1.2, 4.0)	Osogami*	0.538	3	0.1299
	Marie		2	0.0150
	Telek		2	0.0086
Gamma(1.3, 2.0)	Osogami	0.769	2	0.0086
	Marie		2	0.0022
	Telek		2	0.0022
Weibull(3.0, 1.03)	Osogami*	0.943	2	0.0133
	Altiok		2	0.0248
	Telek		2	0.0248
Gamma(0.7, 2.0)	Osogami	1.429	2	0.0248
	Altiok		2	0.1090
	Telek		2	0.1090
Weibull(3.0, 0.6)	Osogami	3.091	2	0.1090
	Altiok		N/A	N/A
	Telek		N/A	N/A
Beta(0.1, 2)	Osogami*	6.452	2	0.7589

*$\alpha_{m+1} \neq 0$

For the purpose of ultimately estimating lifetime distributions, we require that each of the holding time distributions in the semi-Markov process be approximated

with a PH-distribution. This subsection has demonstrated that we can approximate various nonnegative distributions with PH-distributions. While this procedure does increase the size of the original sample space, a newly formed infinitesimal generator matrix, which incorporates these PH-distribution approximations, provides the means to employ the results in Section 3.5 to obtain the lifetime and remaining lifetime distributions.

5.3 Conversion to a Markov Environment

This section details the procedures used to transform a time-homogeneous semi-Markov process (SMP) with generally distributed, nonnegative state holding times, into a time-homogeneous continuous time Markov chain (CTMC) for the estimation of the lifetime distribution. We first summarize the method and then make two lifetime distribution comparisons by using a newly formed CTMC in Equation (3.51) with known $\boldsymbol{\alpha}, \mathbf{R}_D$, and K. These lifetime comparisons provide sufficient evidence to select Osogami and Harchol-Balter's [70] PH-distribution approximation technique as the preferred method for this conversion. We then conclude our analysis by showing that we may use this approach, via the estimation techniques provided in Section 3.5, to compute the lifetime distribution using only degradation data.

5.3.1 Transformation Process

Phase-type distribution approximations to general distributions arise regularly in queueing analyses [58,65,93]. However, methods to transform a SMP into a CTMC using PH-distribution approximations to the general holding time distributions in a SMP are less well known. The holding time distributions in a SMP are the diagonal elements of the semi-Markov kernel, as defined by Equation (4.4), and these diagonal elements are required for a PH-distribution approximation. The off-diagonal elements in the semi-Markov kernel provide the transition probabilities from the current state to the next state. Thus, if we are transforming a SMP into a CTMC, we

require a PH-distribution approximation for the sojourn time distribution of each state as well as the conditional, one-step transition probabilities. Our technique, as graphically depicted in Figure 5.14, incorporates a PH-distribution approximation to each holding time distribution in the semi-Markov process as well as either Equations (3.72 - 3.73) or Equations (3.75 - 3.76). These equations provide the method to transition from the absorbing phase of the current PH-distribution approximation to the initial phase of the next PH-distribution approximation. Each $\hat{q}_{i,j}$ is multiplied by some large value, M, in order to make this transition near instantaneous.

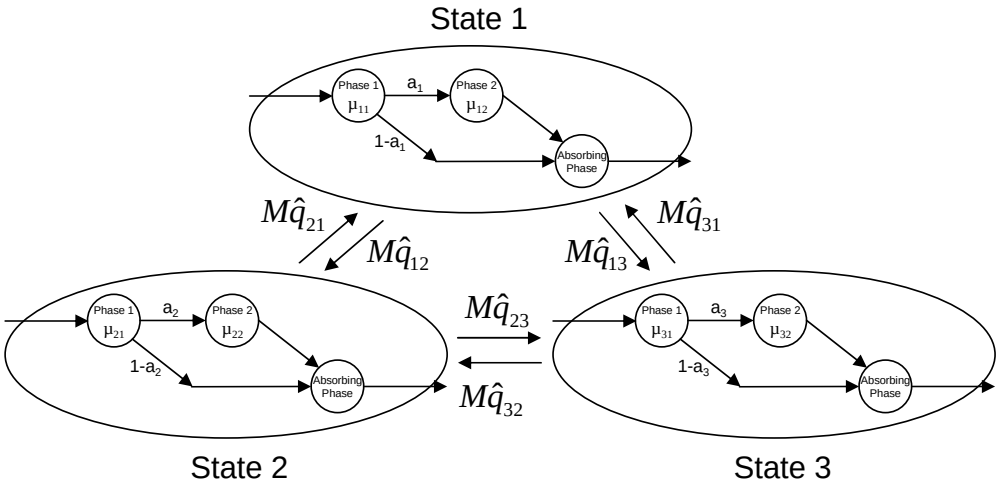

Figure 5.14: Graphical depiction of modified state transition diagram ($K = 3$)

Mathematically, a three-state SMP with general holding time distributions, $H_1(w), H_2(w)$, and $H_3(w)$, and PH-distribution approximations

$$\hat{H}_1(w) = 1 - \boldsymbol{\alpha}_1 e^{\mathbf{T}_1 w} \boldsymbol{e}_1$$
$$\hat{H}_2(w) = 1 - \boldsymbol{\alpha}_2 e^{\mathbf{T}_2 w} \boldsymbol{e}_2$$
$$\hat{H}_3(w) = 1 - \boldsymbol{\alpha}_3 e^{\mathbf{T}_3 w} \boldsymbol{e}_3$$

with associated $\mathbf{T}_1^0, \mathbf{T}_2^0$, and $\mathbf{T}_3^0$ would result in the expanded infinitesimal generator matrix,

$$\hat{\mathbf{Q}}_s = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_1^0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & M\hat{q}_{1,1} & \mathbf{M}_{1,2} & 0 & \mathbf{M}_{1,3} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_2 & \mathbf{T}_2^0 & \mathbf{0} & \mathbf{0} \\ \mathbf{M}_{2,1} & 0 & 0 & M\hat{q}_{2,2} & \mathbf{M}_{2,3} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{T}_3 & \mathbf{T}_3^0 \\ \mathbf{M}_{3,1} & 0 & \mathbf{M}_{3,2} & 0 & 0 & M\hat{q}_{3,3} \end{bmatrix} \tag{5.39}$$

where $\mathbf{T}_i$ is a $m_i \times m_i$ matrix, $\mathbf{T}_i^0$ is a $m_i \times 1$ column vector, and $\mathbf{M}_{i,j}$ is a $1 \times m_j$ row vector where $\mathbf{M}_{i,j} = [M\hat{q}_{i,j} \quad 0 \quad \dots \quad 0]$. Through this representation, we can immediately see a major disadvantage of this approach. The three-state SMP will expand to at least a nine-state CTMC because each $\hat{H}_i(w)$ has, at a minimum, a two-phase PH-distribution approximation and an absorption state. Of course, this assumes that the general distribution does not have $c^2 = 1$ which indicates $H_i(w)$ is an exponential distribution requiring only one phase.

Using this information, we next form lifetime distributions by simulating an SMP environment and then comparing lifetime distributions with those formed via Equation (3.51) where we assume $\boldsymbol{\alpha}, \mathbf{R}_D$, and K are known and that $\mathbf{Q} = \hat{\mathbf{Q}}_s$ as shown in Equation (5.39). Our purpose is to demonstrate that we can adequately transform the SMP into a CTMC while at the same time comparing Marie's [58], Altiok's [5], Telek and Heindl's [86] and Osogami and Harchol-Balter's [70] PH-distribution approximation techniques.

5.3.2 Lifetime Distribution Comparisons

In this section, we compare two sets of lifetime distributions. The first set compares the lifetime distribution from a simulated three-state SMP environment, a twelve-state CTMC environment formed from Telek and Heindl's [86], Marie's [58],

and Altiok's [5] PH-distribution approximations, and a thirteen-state CTMC environment formed from Osogami and Harchol-Balter's [70] PH-distribution approximations. The second set compares the lifetime distribution from a simulated three-state SMP environment with a sixteen-state CTMC environment formed from Osogami and Harchol-Balter's [70] PH-distribution approximations.

The first set of comparisons has an SMP environment where the holding times are distributed according to the following distributions,

$$\begin{aligned} H_1(w) &\sim \text{Gamma}(1.3, 2.0) \\ H_2(w) &\sim \text{Beta}(2.0, 4.0) \\ H_3(w) &\sim \text{Weibull}(3.0, 0.6). \end{aligned}$$

with degradation rates,

$$\mathbf{R}_D = \begin{bmatrix} 0.4 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 4.0 \end{bmatrix}$$

and transition probability matrix,

$$\mathbf{P} = \begin{bmatrix} 0 & 0.6 & 0.4 \\ 0.1 & 0 & 0.9 \\ 0.25 & 0.75 & 0 \end{bmatrix}.$$

The twelve-state CTMC environment, with infinitesimal generator, $\hat{\mathbf{Q}}_s$, uses the lowest MAD in probability of the non-Osogami PH-distribution approximations for these general distributions. We used Telek and Heindl's [86] PH-distribution approximation for Gamma(1.3, 2.0), Marie's [58] PH-distribution approximation for Beta(2.0, 4.0) and Altiok's [5] PH-distribution approximation for Weibull(3.0, 0.6)

where $\mathbf{T}_1$ is given by Equation (5.27), $\mathbf{T}_2$ is given by Equation (5.20) and $\mathbf{T}_3$ is given by Equation (5.35). Additionally, for the transition rates, $q_{i,j}$,

$$\mathbf{Q} = \begin{bmatrix} -1.0 & 0.6 & 0.4 \\ 0.1 & -1.0 & 0.9 \\ 0.25 & 0.75 & -1 \end{bmatrix}$$

and $M = 100000$. Since Telek and Heindl's [86] PH-distribution approximation has $\alpha_1 \neq 1$, we must add an additional state to $\hat{\mathbf{Q}}_s$. We modify $\mathbf{T}_1$, using $M = 100000$ to make the transition near-instantaneous, to include this initial probability distribution where

$$\mathbf{T}_1 = \begin{bmatrix} -100000 & 78000 & 22000 \\ 0 & -0.47 & 0.47 \\ 0 & 0 & -1.07 \end{bmatrix}, \tag{5.40}$$

in place of

$$\mathbf{T}_1 = \begin{bmatrix} -0.47 & 0.47 \\ 0 & -1.07 \end{bmatrix}. \tag{5.41}$$

Additionally, we must adjust the degradation rate matrix, $\mathbf{R}_D$, for the newly formed CTMC. For this twelve-state CTMC,

$$\mathbf{R}_D(i,j) = \begin{cases} 0.4, & i = j, \quad 1 \leq i \leq 4 \\ 1.5, & i = j, \quad 5 \leq i \leq 9 \\ 4.0, & i = j, \quad 10 \leq i \leq 12 \\ 0, & \text{elsewhere} \end{cases}. \tag{5.42}$$

The third lifetime distribution in this first set of comparisons uses a thirteen-state CTMC environment using Osogami and Harchol-Balter's [70] PH-distribution

approximations for gamma(1.3, 2.0), beta(2.0, 4.0), and Weibull(3.0, 0.6) distributions where $\mathbf{T}_1$ is given by Equation (5.28), $\mathbf{T}_2$ is given by Equation (5.21), and $\mathbf{T}_3$ is given by Equation (5.37). The generator matrix, $\mathbf{Q}$, and M-value are the same as the twelve-state CTMC. For the matrix, $\mathbf{T}_2$, we append another row because $\boldsymbol{\alpha}$ has $\alpha_1 \neq 1$. In this case, both $\mathbf{T}_2$ and $\mathbf{T}_2^0$ require modification using $M = 100000$ resulting in

$$\mathbf{T}_2 = \begin{bmatrix} -100000 & 93300 & 0 & 0 & 0 & 0 \\ 0 & -14 & 14 & 0 & 0 & 0 \\ 0 & 0 & -14 & 14 & 0 & 0 \\ 0 & 0 & 0 & -14 & 14 & 0 \\ 0 & 0 & 0 & 0 & -14 & 14 \\ 0 & 0 & 0 & 0 & 0 & -14 \end{bmatrix} \tag{5.43}$$

and

$$\mathbf{T}_2^0 = \begin{bmatrix} 6700 \\ 0 \\ 0 \\ 0 \\ 0 \\ 14 \end{bmatrix}. \tag{5.44}$$

For this thirteen-state CTMC,

$$\mathbf{R}_D(i,j) = \begin{cases} 0.4, & i = j, \quad 1 \leq i \leq 3 \\ 1.5, & i = j, \quad 4 \leq i \leq 10 \\ 4.0, & i = j, \quad 11 \leq i \leq 13 \\ 0, & \text{elsewhere} \end{cases}. \tag{5.45}$$

The expanded generator matrix, $\hat{\mathbf{Q}}_s$, associated initial probability distribution, $\boldsymbol{\alpha}$, and $\hat{\mathbf{R}}_D$ are used to obtain the lifetime distribution via

$$\widetilde{F}(u,w) = 1 - \boldsymbol{\alpha}\exp((\hat{\mathbf{Q}}_s - u\hat{\mathbf{R}}_D)w)\boldsymbol{e}, \quad Re(u) > 0,$$

and the remaining lifetime distribution via

$$\hat{R}(x,w|\xi_0) = \frac{1-\hat{F}(x,w+\xi_0)}{1-\hat{F}(x,\xi_0)}.$$

Comparisons between the three lifetime and remaining lifetime distributions at failure threshold level $x = 50$ are shown in Figure 5.15 and Table 5.12 where $\xi_0 = 42.194$.

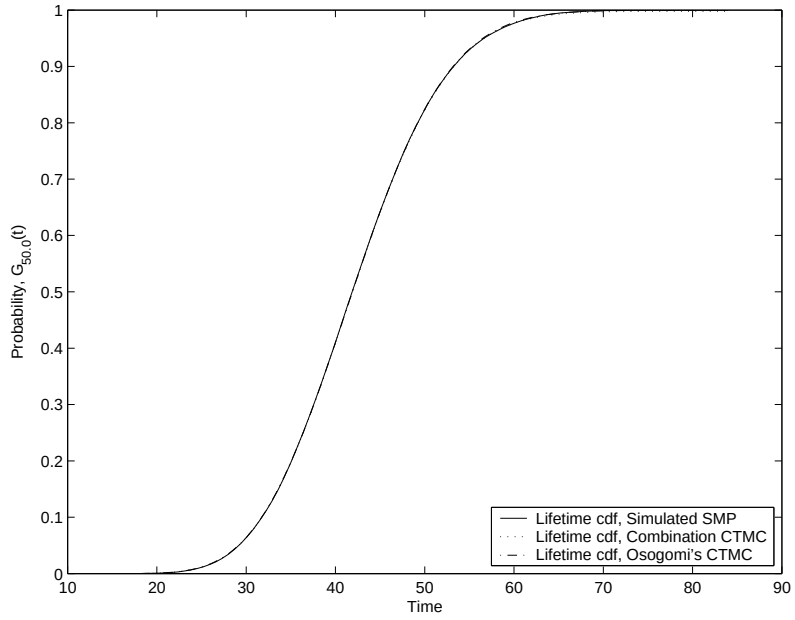

Figure 5.15: CDF comparisons for 3-state SMP, 12-state CTMC, and 13-state CTMC.

Table 5.12: Cramér-von Mises test statistics: PH-distribution approximation techniques. (κ^*=0.461, $\alpha = 0.05$).

Technique	Number of States	$\hat{F}(x,w)$	$1-\hat{R}(x,w\|\xi_0)$
Combination	12	9.412E-05	5.632E-04
Osogami and Harchol-Balter	13	2.605E-05	1.157E-04

In this example, Osogami and Harchol-Balter [70] provide a better approximation than Telek and Heindl's [86], Marie's [58], and Altiok's [5] combined technique for the lifetime distribution and the remaining lifetime distribution.

We provide one other set of comparisons based on a three-state SMP where

$$\begin{aligned} H_1(w) &\sim \text{Beta}(2.0, 4.0) \\ H_2(w) &\sim \text{Beta}(0.1, 2.0) \\ H_3(w) &\sim \text{Beta}(1.2, 4.0) \end{aligned}$$

with degradation rates,

$$\mathbf{R}_D = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & 2.0 \end{bmatrix}$$

and transition probability matrix,

$$\mathbf{P} = \begin{bmatrix} 0 & 0.7 & 0.3 \\ 0.6 & 0 & 0.4 \\ 0.2 & 0.8 & 0 \end{bmatrix}.$$

The PH-distribution approximations have $\mathbf{T}_1$ as defined by Equation (5.21), $\mathbf{T}_2$ as defined by Equation (5.38), and $\mathbf{T}_3$ as defined by Equation (5.25), but each $\mathbf{T}_i$ and $\mathbf{T}_i^0$ were modified similarly to that shown in Equations (5.43) and (5.44) because the initial probability distribution is not $[1 \quad 0 \quad \dots \quad 0]$. In this comparison,

$$\mathbf{Q} = \begin{bmatrix} -1.0 & 0.7 & 0.3 \\ 0.6 & -1.0 & 0.4 \\ 0.2 & 0.8 & -1 \end{bmatrix}$$

and $M = 100000$ resulting in a 16×16 generator matrix for this CTMC environment. The lifetime distributions are shown in Figure 5.16 where the Cramér-von Mises test statistic for $\hat{F}(x, w)$ is 2.455E-05 and for $1 - \hat{R}(x, w|\xi_0)$, at $\xi_0 = 12.464$, is 5.358E-05. No combination of the other approximation techniques would produce a lifetime or remaining lifetime distribution.

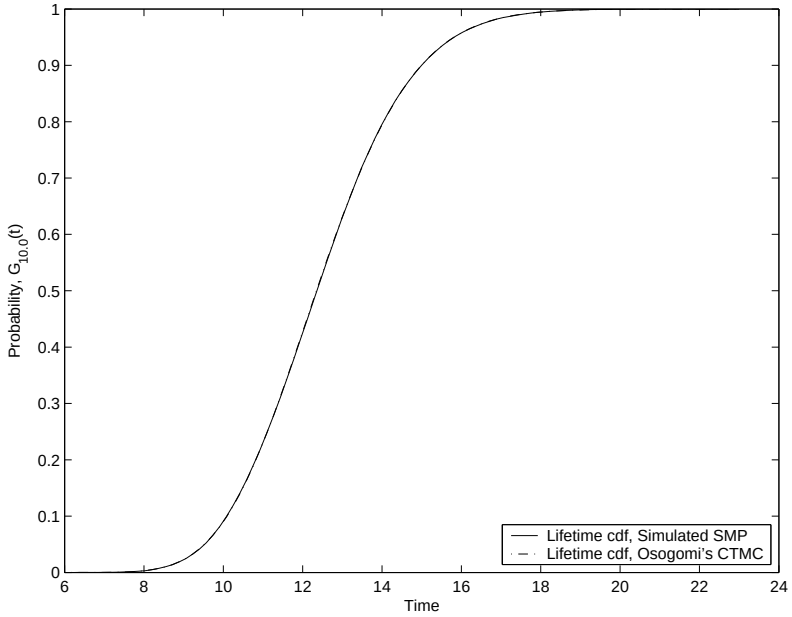

Figure 5.16: CDF comparisons for 3-state SMP and 16-state CTMC environments.

These two sets of comparisons have addressed two important issues. The first is whether the lifetime distribution resulting from the transformation from an SMP environment to a CTMC environment retains enough accuracy to be practically valuable. The maximum absolute deviation for both sets of comparisons were less than 0.003 in probability which indicates the estimate is accurate in these cases. The second issue is the selection of the best PH-distribution approximation technique from the combination of Marie [58], Altiok [5], Telek and Heindl [86], and Osogami and Harchol-Balter [70]. While Osogami and Harchol-Balter's technique might add an extra state for each PH-distribution approximation, their approximation was closer in the first set of comparisons and the other PH-distribution approximation techniques could not be utilized to form a lifetime distribution in the second set of

comparisons. Thus, it is concluded that Osogami and Harchol-Balter's technique best suits our purposes.

While it is possible to adequately approximate the lifetime distribution by transforming the SMP environment into a CTMC environment, it is not clear that we can effectively estimate the initial probability distribution, the degradation rates, the transition probabilities, the number of environment states, and the PH-distribution approximations given only a sequence of degradation observations. The next subsection demonstrates that we can, in fact, perform these operations using only degradation data.

5.3.3 Observable Degradation Comparisons

This section demonstrates that the numerical methods developed in Chapter III to find the lifetime distribution for a single-unit system subject to a CTMC environment, coupled with the methods presented in this chapter to transform a SMP into a CTMC, present a viable method to obtain the lifetime distribution (and hence the remaining lifetime distribution) for use in systems prognosis. While we rely on Equation (3.51), once the degradation rates are estimated, $\hat{\mathbf{R}}_D$ is modified in a manner similar to that shown in Equations (5.42) and (5.45). Additionally, we let $\mathbf{Q} = \hat{\mathbf{Q}}_s$. To complete the numerical implementation procedures, Osogami and Harchol-Balter [70] PH-distribution approximation required the first three noncentral moments, μ_1, μ_2, and μ_3, and those are estimated by [91]

$$\hat{\mu}_1 = \frac{1}{n_s}\sum_{i=1}^{n_s} Y_i, \quad \hat{\mu}_2 = \frac{1}{n_s}\sum_{i=1}^{n_s} Y_i^2, \quad \hat{\mu}_3 = \frac{1}{n_s}\sum_{i=1}^{n_s} Y_i^3 \tag{5.46}$$

where n_s is the total number of holding time observations and Y_i is the ith holding time observation.

We generated 1000 sample degradation paths using a five-state SMP with respective state sojourn distributions

$$\begin{aligned} H_1(w) &\sim \text{Gamma}(0.7, 2.0) \\ H_2(w) &\sim \text{Weibull}(3.0, 1.03) \\ H_3(w) &\sim \text{Gamma}(2.5, 2.0) \\ H_4(w) &\sim \text{Weibull}(2.5, 3.0) \\ H_5(w) &\sim \text{Beta}(0.1, 2.0) \end{aligned}$$

and associated state degradation rates,

$$\mathbf{R}_D = \begin{bmatrix} 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0.75 & 0 & 0 & 0 \\ 0 & 0 & 1.1 & 0 & 0 \\ 0 & 0 & 0 & 1.85 & 0 \\ 0 & 0 & 0 & 0 & 2.4 \end{bmatrix}$$

with transition probability matrix,

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.3 & 0 & 0.1 & 0.4 \\ 0.1 & 0.1 & 0.2 & 0 & 0.6 \\ 0.3 & 0.2 & 0.2 & 0.3 & 0 \end{bmatrix}.$$

For each of the 1000 sample degradation paths, we implemented the procedure of Section 3.5.1, where at $t_0 = 0$, $\hat{X}(t_0) = 0$. We observed the degradation paths until $t_{200} = 300$ and then approximated these sample paths by creating piecewise-linear approximations. From these new sample paths, we collected all of the approximated rates and performed K-means cluster analysis for $K = 2, 3, \ldots, 15$. At this point, we

calculated the F_K-values shown in Table 5.13 which indicates the first local maximum at $\hat{K} = 3$. For completeness, we also examine $\hat{K} = 2$ and $\hat{K} = 5$.

Table 5.13: Values of $F_K(\times 10^6)$

K	F_K
2	0.0712
3	0.7848
4	0.6457
5	1.1511
6	1.1467
7	1.7381
8	1.6264
9	1.5232
10	2.1041
11	2.7835
12	2.9877
13	2.9603
14	1.3027
15	3.5333

The estimated degradation rates for $\hat{K} = 2, 3,$ and 5 are given in Table 5.14. We then use these new degradation rates to form new sample paths and find $\hat{q}_{i,j}$.

Table 5.14: Estimated degradation rates.

$\hat{r}(i)$	$\hat{K} = 2$	$\hat{K} = 3$	$\hat{K} = 5$
1	1.1317	1.5092	0.2582
2	0.3581	0.3404	1.7089
3	—	1.0717	1.3616
4	—	—	1.0895
5	—	—	0.7030

For $\hat{K} = 2$,

$$\hat{\mathbf{Q}} = \begin{bmatrix} -0.0875 & 0.0875 \\ 0.2939 & -0.2939 \end{bmatrix}, \tag{5.47}$$

for $\hat{K} = 3$,

$$\hat{\mathbf{Q}} = \begin{bmatrix} -0.5459 & 0.1044 & 0.4415 \\ 0.0567 & -0.2985 & 0.2418 \\ 0.0611 & 0.0787 & -0.1398 \end{bmatrix}, \tag{5.48}$$

and for $\hat{K} = 5$,

$$\hat{\mathbf{Q}} = \begin{bmatrix} -0.3192 & 0.0163 & 0.0432 & 0.1381 & 0.1216 \\ 0.0725 & -0.5973 & 0.1162 & 0.3285 & 0.0801 \\ 0.0665 & 0.0451 & -0.5802 & 0.4016 & 0.0670 \\ 0.0384 & 0.0183 & 0.0526 & -0.1661 & 0.0568 \\ 0.1820 & 0.0309 & 0.0637 & 0.2970 & -0.5736 \end{bmatrix}. \tag{5.49}$$

Since $H^{(k)}(i)$ in Equation (3.76) is the sum of the individual holding times, we can easily collect these holding times and use Equation (5.46) to estimate the moments for Osogami and Harchol-Balter's [70] PH-distribution approximation technique. For $\hat{K} = 2$, $\hat{\mathbf{Q}}_s$ is an 8×8 matrix where

$$\mathbf{T}_1 = \begin{bmatrix} -0.1051 & 0.1051 & 0 \\ 0 & -1.0774 & 0.1155 \\ 0 & 0 & -0.1080 \end{bmatrix}$$

and

$$\mathbf{T}_2 = \begin{bmatrix} -0.5320 & 0.5320 & 0 \\ 0 & -0.6775 & 0.0043 \\ 0 & 0 & -0.1362 \end{bmatrix}.$$

For $\hat{K} = 3$, $\hat{\mathbf{Q}}_s$ is a 17×17 matrix where

$$\mathbf{T}_1 = \begin{bmatrix} -3.33 & 3.33 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3.33 & 3.33 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3.33 & 3.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3.33 & 3.33 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3.33 & 3.33 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -3.33 & 3.33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -33.655 & 0.00016 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.0345 \end{bmatrix},$$

$$\mathbf{T}_2 = \begin{bmatrix} -0.5422 & 0.5422 & 0 \\ 0 & -0.6838 & 0.0039 \\ 0 & 0 & -0.1346 \end{bmatrix},$$

and

$$\mathbf{T}_3 = \begin{bmatrix} -0.1899 & 0.1899 & 0 \\ 0 & -0.9586 & 0.1427 \\ 0 & 0 & -0.1765 \end{bmatrix}.$$

Finally, for $\hat{K} = 5$, $\hat{\mathbf{Q}}_s$ is a 42×42 matrix where

$$\mathbf{T}_1 = \begin{bmatrix} -0.6047 & 0.6047 & 0 \\ 0 & -0.6823 & 0.00085 \\ 0 & 0 & -0.0916 \end{bmatrix},$$

$$\mathbf{T}_2(i,j) = \begin{cases} -5.9871, & i = j, \qquad 1 \le i \le 10 \\ 5.9871, & j = i+1, \qquad 1 \le i \le 10 \\ -256.35, & i = j = 11 \\ 1.9998 \times 10^{-6}, & i = 11, j = 12 \\ -0.00495, & i = j = 12 \\ 0, & \text{elsewhere} \end{cases},$$

$$\mathbf{T}_3(i,j) = \begin{cases} -4.9387, & i = j, \qquad 1 \le i \le 8 \\ 4.9387, & j = i+1, \qquad 1 \le i \le 8 \\ -9.6513, & i = j = 9 \\ 2.6148 \times 10^{-4}, & i = 9, j = 10 \\ -0.07246, & i = j = 10 \\ 0, & \text{elsewhere} \end{cases},$$

$$\mathbf{T}_4 = \begin{bmatrix} -0.2482 & 0.2482 & 0 \\ 0 & -0.7784 & 0.1131 \\ 0 & 0 & -0.2057 \end{bmatrix},$$

and

$$\mathbf{T}_5(i,j) = \begin{cases} -4.3293, & i = j, \qquad 1 \le i \le 7 \\ 4.3293, & j = i+1, \qquad 1 \le i \le 7 \\ -7.9305, & i = j = 8 \\ 2.6284 \times 10^{-4}, & i = 8, j = 9 \\ -0.0705, & i = j = 9 \\ 0, & \text{elsewhere} \end{cases}.$$

Lastly, for $\hat{K} = 2$, $\hat{\boldsymbol{\alpha}}$ is a 1×8 row vector where $\hat{\alpha}_1 = 0.335$, $\hat{\alpha}_5 = 0.665$ and $\hat{\alpha}_i = 0$, elsewhere. For $\hat{K} = 3$, $\hat{\boldsymbol{\alpha}}$ is a 1×17 row vector where $\hat{\alpha}_1 = 0.075$, $\hat{\alpha}_{10} = 0.649$, $\hat{\alpha}_{14} = 0.276$,and $\hat{\alpha}_i = 0$, elsewhere. For $\hat{K} = 5$, $\hat{\boldsymbol{\alpha}}$ is a 1×42 row vector where $\hat{\alpha}_1 = 0.531$, $\hat{\alpha}_5 = 0.016$, $\hat{\alpha}_{18} = 0.078$, $\hat{\alpha}_{29} = 0.171$, $\hat{\alpha}_{33} = 0.204$, and $\hat{\alpha}_i = 0$, elsewhere. Using the estimates for $\boldsymbol{\alpha}, \mathbf{R}_D$, and $\mathbf{Q}$ in Equation (3.51) for $\hat{K} = 2, 3$, and 5, we obtain the lifetime distributions shown in Figure 5.17 which are compared to the simulated SMP at failure threshold level $x = 400$. The numerical comparisons are provided in Table 5.15, $\xi_0 = 420.118$, where we fail to reject the null hypothesis that the simulated distribution and the analytical distributions are equivalent at $\alpha = 0.05$. Thus, we conclude that it is possible to obtain the lifetime and remaining lifetime distribution of a single-unit system subject to a semi-Markov environment.

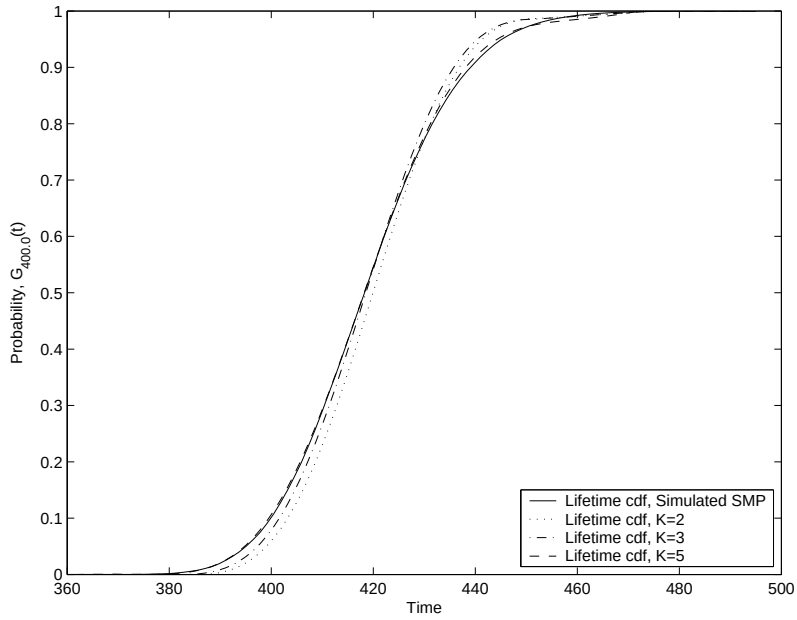

Figure 5.17: Lifetime comparisons of a system subject to a SMP environment.

This chapter demonstrates a methodology that removed the requirement to utilize the difficult to implement, analytical lifetime distributions derived in Chapter IV. By converting the semi-Markovian environment into a Markovian environment, we utilize the analytical lifetime distribution derived in Chapter III and show that

Table 5.15: Cramér-von Mises statistic for SMP conversion (κ^*=0.461, $\alpha = 0.05$).

$\hat{K}$	Number of States	$\hat{F}(x,w)$	$1-\hat{R}(x,w\|\xi_0)$
2	8	3.172E-02	5.331E-02
3	17	1.314E-02	7.284E-02
5	42	7.649E-04	5.061E-03

both observable environment and observable degradation data provide the necessary holding time observations to approximate the PH-distributions required to make this conversion. This capability includes and extends the applicability of the results contained in Chapter III. If sensors, attached to a component, provide information on the observable environment or on the observable degradation, then assuming the environment can be modelled as a homogeneous semi-Markovian process, we can obtain the remaining lifetime distribution and the failure time probabilities required to make probabilistic decisions that impact autonomic logistics. In the next chapter, we conclude by providing our contributions, recommendations, and future research.

VI. Conclusions, Recommendations, and Future Work

This chapter provides a summary of the methodological and practical contributions of this dissertation in addition to recommendations for the actual implementation of these procedures. Lastly, we provide areas of future research interest.

6.1 Dissertation Contributions

The first methodological contribution was the mathematical characterization of a single-unit system's randomly evolving environment. We characterized this environment as a nonhomogeneous, continuous-time Markov chain (CTMC), a homogeneous CTMC, and a homogeneous, continuous-time semi-Markov process (SMP). For the nonhomogeneous CTMC environment, we obtained an analytical result for the lifetime distribution of the system and demonstrated that we can obtain the moments of this distribution in a simple case. For the homogeneous CTMC environment, we obtained an analytical result for the lifetime distribution of the system and demonstrated that we can obtain the moments of this distribution in all cases. For the homogeneous SMP environment, we obtained an analytical result for the lifetime distribution of the system.

The practical contributions involve the numerical implementation of the analytical results. Assuming a homogeneous CTMC environment process, we provide statistical estimation techniques for an observable environment and observable degradation. This numerical implementation imparts the remaining lifetime distribution in both scenarios and illustrates that, if a sensor provides environment or degradation data, then systems prognosis is feasible. For the case of a homogeneous SMP environment, we provided approximation procedures via phase-type (PH) distributions. This numerical implementation also provides the remaining lifetime distribution and

greatly expands its applicability to an observable environment and observable degradation with non-exponential environment state holding times. We next offer some recommendations on the implementation of these procedures.

6.2 Recommendations

The practical contributions specifically mentioned two scenarios: observable environment and observable degradation. If sensors provide data on the environment, then we recommend estimating the squared coefficient of variation for the holding times in each state. If this value is close to one, then the procedures developed for the homogeneous CTMC environment developed in Sections 3.2.2 and 3.5 are applicable; otherwise, procedures developed for the homogeneous SMP environment detailed in Section 5.3 are required. If sensors provide degradation data, the number of states and degradation rates must be estimated first. With this information, the squared coefficient of variation for the holding times should be determined and the recommendation for the observable environment followed. However, if there are concerns that the time-homogeneity assumption is violated, these techniques should not be implemented.

Converting the SMP to a CTMC is a challenging process. While we recommend using Osogami and Harchol-Balter's [70] technique in most cases, if the squared coefficient for each distribution is less than 0.5, then we recommend using Marie's [58] technique. PH-distributions are extremely powerful, but methods to construct these approximations vary significantly. Thus, practitioners should take care to implement the appropriate approximation for a given system.

6.3 Future Research

Over the course of this research, potentially fruitful areas for future research were discovered. We present those that hold the most promise.

In order to solve the partial differential equation leading to the lifetime distribution, we were required to take the Laplace transform of the PDE; However, weak analytical solutions may exist. Additionally, it may be possible to obtain an approximate numerical solution and bypass the Laplace transform.

While we were unable to numerically implement the non-homogeneous environment, it is possible that an age- or degradation-state dependent model may be more appropriate and provide the means to determine the lifetime distribution numerically. As an example, consider the analysis of the Virkler [90] data. The analysis may have failed due to the fact that we applied a time-stationary approach to a possibly non-stationary data set.

A better method to estimate the number of environment states should be developed. In a real-time environment, our procedure may become slow, resulting in delayed results. It was noticed that a two- or three-state CTMC often provided better results for the lifetime distribution than the estimate $\hat{K}$. While it is possible that this observation holds only in the cases we examined, additional study of this phenomenon could be very fruitful as the benefits of a smaller state size are enormous.

This dissertation examined scenarios having a small number of states. If the state size increased dramatically, we may be unable to estimate the lifetime distribution in a timely manner. Numerical experiments indicated that as the number of states increased, the time required to perform the K-means clustering technique also increased. Additionally, numerical inversion of Laplace transforms becomes more costly (and error prone) as the state size increases. Methods to account for a large number of states should be explored.

Lastly, no exploration of error propagation was conducted. Error is compounded possibly in the estimation of the parameters for the generator matrix, the degradation rates and the number of environment states. As the number of environment states increases, so does the number of parameters to be estimated and the error associated with the model. Better estimation techniques and an associated

minimal number of environment states would help alleviate this problem. Once the estimates are incorporated in the lifetime distribution, an inverse transform must be performed. This inverse transform result introduces additional error because it is an approximation of the actual lifetime distribution. Obtaining the exact lifetime distribution analytically or possibly an approximate numerical solution to the PDE in lieu of numerical Laplace transform inversion would also help.

The United States Air Force is developing its next generation aircraft and is seeking to reduce the risk of catastrophic failures, maintenance activities, and the logistics footprint while improving its sortie generation rate through a process called autonomic logistics. Vital to the successful implementation of this process is remaining lifetime prognosis of critical aircraft components. Complicating this problem is the absence of failure time information; however, sensors located on the aircraft are providing degradation measures. This research has provided a method to address at least a portion of this problem by uniting analytical lifetime distribution models with environment and/or degradation measures to obtain the remaining lifetime distribution.

Bibliography

1. Abate, J. and Whitt, W. (1995), Numerical inversion of Laplace transforms of probability distributions, *ORSA Journal on Computing*, **7**, 36-43.
2. Abdel-Hameed, M. (1984), Life distribution properties of devices subject to a Lévy wear process, *Mathematics of Operations Research*, **9** (4), 606-614.
3. Albert, A. (1962), Estimating the infinitesimal generator of a continuous time, finite state Markov process, *Annals of Mathematical Statistics*, **33**, 727-753.
4. Aldous, D. and L. Shepp (1987), The least variable phase type distribution is Erlang, *Communications in Statistics-Stochastic Models*, **3** (3), 467-473.
5. Altiok, T. (1985), On the phase-type approximations of general distributions, *IIE Transactions*, **17** (2), 110-116.
6. Asmussen, S. and M. Bladt (1997), Renewal theory and queueing algorithms for matrix-exponential distributions, in *Matrix-Analytic Methods in Stochastic Models*, S.R. Chakravarthy and A.S. Alfa (Eds.), Marcel Dekker, Inc., New York, 313-341.
7. Basawa, I.V. and Prakasa Rao, B.L.S. (1980), *Statistical Inference for Stochastic Processes*, John Wiley and Sons.
8. BestFit Version 4.5 (2002), Computer software, Palisade Corporation, 31 Dekker Road, Newfield, New York 14867.
9. Bladt, M. and M.F. Neuts (2003), Matrix-exponential distributions: Calculus and interpretations via flows, *Stochastic Models*, **19** (1), 113-124.
10. Bogdanoff, J.L. and F. Kozin (1985), *Probabilistic Models of Cumulative Damage*, John Wiley and Sons, New York.
11. *BorgWarner Cooling Systems* (2000), Visctronic^{TM} Direct Actuated Fan Drives brochure.
12. Box, G.E.P., G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*. Revised Ed., Prentice Hall, NJ.
13. "Building a Global Future," http://web.mit.edu/ctl/www/news/48/caterpillar.html.
14. Calinski, R.B. and J. Harabasz (1974), A dendrite method for cluster analysis, *Communications in Statistics*, **3**, 1-27.
15. Chan, V. and W.Q. Meeker, Estimation of degradation-based reliability in outdoor environments. Unpublished manuscript, Iowa State University.
16. Chao, M. (1999), Degradation analysis and related topics: Some thoughts and a review, *Proceedings of the National Science Council ROC(A)*, **23** (5), 555-565.

17. Chatfield, C. (1996), *The Analysis of Time Series, An Introduction*, 5^{th} Ed., Chapman and Hall.

18. Çinlar, E. (1977), Shock and wear models and Markov-additive processes, in *The Theory and Applications of Reliability* (I.N. Shimi and C.P. Tsokos, eds.), Academic, 193-214.

19. Conover, W.J. (1980), *Practical Nonparametric Statistics*. 2^{nd} Ed., John Wiley and Sons, New York.

20. *Cummins Inc.* (2000), QST30 Industrial Specifications brochure.

21. Cumani, A. (1982), On the canonical representation of homogeneous Markov processes modelling failure-time distributions, *Microelectronics and Reliability*, **22**, 583-602.

22. Department of the Army (1996), *Army Material Maintenance Policy and Retail Maintenance Operations* AR 750-1, Washington: HQ USA.

23. Elsayed, E. A. (1996), *Reliability Engineering*. Addison Wesley Longman, Inc., Massachusetts.

24. Esary, J.D., A.W. Marshall, and F. Proschan (1973), Shock models and wear processes, *The Annals of Probability*, **1** (4), 627-649.

25. ExpertFit Distribution Fitting Software, Computer software, Averill M. Law and Associates, Inc., P.O. Box 40996, Tucson, AZ 85717.

26. Ghonem, H. and S. Dore (1987), Experimental study of the constant probability crack growth curves under constant amplitude loading, *Engineering Fracture Mechanics*, **27**, 1-25.

27. Gottlieb, G. (1980), Failure distributions of shock models, *Journal of Applied Probability*, **17**, 745-752.

28. Greitzer, F.L., L.J. Kangas, K.M. Terrones, M.A. Maynard, B.W. Wilson, R.A. Pawlowski, D.R. Sisk, and N.B. Brown (1999), Gas Turbine Engine Health Monitoring and Prognostics, *International Society of Logistics (SOLE) Symposium*, Las Vegas, Nevada.

29. Greitzer, F.L., E.J. Stahlman, T.A. Ferryman, B.W. Wilson, L.J. Kangas, and D.R. Sisk (1999), Development of a Framework for Predicting Life of Mechanical Systems: Life Extension Analysis and Prognostics (LEAP), *International Society of Logistics (SOLE) Symposium*, Las Vegas, Nevada.

30. Hardman, W., A. Hess, R. Ahne, and D. Blunt (2001), USN Development Strategy, Fault Testing Results, and Future Plans for Diagnostics, Prognostics and Health Management of Helicopter Drive Train Systems, *DSTO International Conference on Health and Usage Monitoring (HUMS2001)*, 69-78.

31. Henley, S., R. Currer, B. Scheuren, A. Hess, and G. Goodman (2000), Autonomic logistics - Support concept for the 21st century, *Proceedings of the IEEE Aerospace Conference* 417-421.

32. Hordijk, A. and F.A. Van Der Duyn Schouten (1983), Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model, *Advances in Applied Probability*, **15**, 274-303.

33. Igaki, N., U. Sumita, and M. Kowada (1995), Analysis of Markov renewal shock models, *Journal of Applied Probability*, **32**, 821-831.

34. Jardim-Goncalves, R., M. Martins-Barata, J. Assis-Lopes, and A. Steiger-Garcao (1996), Application of stochastic modelling to support predictive maintenance for industrial environments, *IEEE International Conference on Systems Man and Cybernetics*, **1**, 117-122.

35. Joint Strike Fighter (JSF) website, $http://www.jsf.mil/$

36. Johnson, M.A. (1993), Selecting parameters of phase distributions: Combining nonlinear programming, heuristics, and Erlang distributions, *ORSA Journal on Computing*, **5** (1), 69-83.

37. Johnson, M.A. and M.R. Taaffe (1989), Matching moments to phase distributions: Mixtures of Erlang distributions of common order, *Communications in Statistics-Stochastic Models*, **5**, 711-743.

38. Kangas, L.J., F.L. Greitzer, M. Melius (1997), Automated health monitoring of the M1A1/A2 battle tank, *USAF Inaugural Engine Condition Monitoring (ECM) Workshop*, San Diego, CA.

39. Karlin, S. and H.M. Taylor (1981), *A Second Course in Stochastic Processes*, Academic Press, New York.

40. Karlin, S. and W. Studden (1966), *Tchebycheff Systems: With Applications in Analysis and Statistics*, John Wiley and Sons.

41. Kasumu, R.A. and A Lešanovský (1983), On optimal replacement policy, *Aplikace Matematiky*, **28**, 317-329.

42. Kay, S.M. (1993), *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, New Jersey.

43. Keeney, M.M., R.E. Rhoads, C.D. Taylor (2002), Embedded diagnostics and prognostics synchronization, *Army AL & T*, Sept-Oct, 19-22.

44. Kharoufeh, J.P. (2003), Explicit results for wear processess in a Markovian environment, *Operations Research Letters*, **31** (3), 237-244.

45. Kharoufeh, J.P. and S.M. Cox (2004), Stochastic models for degradation-based reliability, Forthcoming in *IIE Transactions on Quality and Reliability Engineering*.

46. Kharoufeh, J.P. and J. Sipe (2004), Evaluating failure time probabilities for a Markovian wear process, Forthcoming in *Computers and Operations Research*.

47. Kleinrock, L. (1975), *Queueing Systems, Volume 1: Theory*, John Wiley and Sons.

48. Kolesar, P. (1966), Minimum cost replacement under Markovian deterioration, *Management Science*, **12** (9), 694-706.

49. Kreyszig, E. (1988), *Advanced Engineering Mathematics*, John Wiley and Sons, New York.

50. Kulkarni, V.G. (1995), *Modeling and Analysis of Stochastic Systems*, Chapman and Hall.

51. Kulkarni, V., V. Nicola, and K. Trivedi (1986). On modeling the performance and reliability of multi-mode computer systems. *Journal of Systems and Software*, **6**, 175-182.

52. Kulkarni, V.G., V.F. Nicola, and K. Trivedi (1987). The completion time of a job on a multi-mode system. *Advances in Applied Probability*, **19**, 932-954.

53. Lang, A. and J.L. Arthur (1997), Parameter approximation for phase-type distributions, in *Matrix-Analytic Methods in Stochastic Models*, S.R. Chakravarthy and A.S. Alfa (Eds.), Marcel Dekker, Inc., New York, 151-206.

54. Lee, B.H. and S.B. Lee (2000), Stochastic modelling of low-cycle fatigue damage in 316L stainless steel under variable multiaxial loading, *Fatigue Fract Engineering Material Structures*, **23**, 1007-1018.

55. Li, Y., C. Zhang, T.R. Kurfess, S. Dahyluk, and S.Y. Liang (2000), Diagnostics and prognostics of a single surface defect on roller bearings, *Proceedings on the Institution of Mechanical Engineers*, **214 Part C**, 1173-1185.

56. Li, C. and A. Ray (1995), Neural network representation of fatigue damage dynamics, *Fatigue Damage Dynamics* 126-133.

57. Lu, C.J. and W.Q. Meeker (1993), Using degradation measures to estimate a time-to-failure distribution, *Technometrics*, **35** (2), 161-174.

58. Marie, R. (1980), Calculating equilibrium probabilities for $\lambda(n)/C_k/1/N$ queues, *Proceedings of Performance 80*, **9** (2), 117-125.

59. Meeker, W.Q. and L.A. Escobar (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons.

60. Meeker, W.Q., L.A. Escobar, and V. Chan (2002), Using accelerated tests to predict service life in highly variable environments, in *Service Life Prediction: Methodologies and Metrologies*, edited by Bauer, D.R. and Martin, J.W., American Chemical Society, Washington, D.C.

61. Milligan, G.W. and M.C. Cooper (1985), An examination of procedures for determining the number of clusters in a data set, *Psychometrika*, **50** (2), 159-179.

62. Moorthy, M.V. (1995), Numerical inversion of two-dimensional Laplace transforms-Fourier series representation, *Applications in Numerical Mathematics*, **17**, 119-127.

63. National Institute of Child Health and Development (2001), Targeting Sudden Infant Death Syndrome (SIDS): A Strategic Plan.

64. Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman (1996), *Applied Linear Statistical Models*, 4^{th} Ed., Irwin.

65. Neuts, M. F. (1981), *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Dover Publications, Inc., New York.

66. Nolan, M. and T. Cesarone (2002), Implementation of prognostics framework on Navy battle group-automated maintenance environment (BG-AME) and total ship monitoring (TSM) programs, *Proceedings of the IEEE Autotestcon*, 788-800.

67. Oberhettinger, F. and L. Badii (1973), *Tables of Laplace Transforms*, Springer-Verlag, New York.

68. Oluyede, B.O. (2002), Some inequalities and bounds for weighted reliability measures, *Journal of Inequalities in Pure and Applied Mathematics*, **3** (4), Article 60.

69. Olmsted, J.M.H. (1956), *Real Variables: An Introduction to the Theory of Functions*, Appleton-Century-Crofts, Inc., New York.

70. Osogami, T. and M. Harchol-Balter (2003), A closed-form solution for mapping general distributions to minimal PH distributions, *Lecture Notes in Computer Science*, **2794**, 200-217.

71. Park, K, J.Y. Seol, C. Yoo, Y.W. Kim, S.K. Han, E.H. Lee, C. Kim, Y. Shim, and C. Lee (2001), Adenovirus expressing $p27^{Kipl}$ induces growth arrest of lung cancer cell lines and suppresses the growth of established lung cancer xenografts, *Lung Cancer*, **31**, 149-155.

72. Patankar, R. and A. Ray (2000), State-space modeling of fatigue crack growth in ductile alloys, *Engineering Fracture Mechanics*, **66**, 129-151.

73. Perros, H. (1994), *Queueing Networks with Blocking*, Oxford University Press, New York.

74. Puri, P.S and H. Singh (1986), Optimum replacement of a system subject to shocks: A mathematical lemma, *Operations Research*, **34** (5), 782-789.

75. Ray, A. and S. Tangirala (1996), Stochastic modeling of fatigue crack dynamics for on-line failure prognostics, *IEEE Transactions on Control Systems Technology*, **4** (4), 443-451.

76. Ray, A. and S. Tangirala (1997), A nonlinear stochastic model of fatigue crack dynamics, *Probabilistic Engineering Mechanics*, **12** (1), 33-40.

77. Ray, A. (1999), Stochastic modeling of fatigue crack damage for risk analysis and remaining life prediction, *Journal of Dynamic Systems, Measurement, and Control*, **121**, 386-393.

78. Ray, A. and R. Patankar (1999), A stochastic model of fatigue crack propagation under variable-amplitude loading, *Engineering Fracture Mechanics*, **62**, 477-493.

79. Sauer, C. and K. Chandy (1975), Approximate analysis of central server models, *IBM Journal of Research and Development*, **19**, 301-313.

80. Shanthikumar, J.G. and U. Sumita (1983), General shock models associated with correlated renewal sequences, *Journal of Applied Probability*, **20**, 600-614.

81. Smith, G., J.B. Schroeder, and B.L. Masquelier (1999), Logistics for the joint strike fighter-It ain't business as usual, *Air Force Journal of Logistics*, **23** (1), 13-17.

82. Su, L.P., M. Nolan, and D.R. Carey (1998), Prognostics framework for weapon systems health management, *Proceedings of the IEEE Autotestcon.*

83. Su, L.P., M. Nolan, G. deMare, and D.R. Carey (1999), Prognostics framework, *Proceedings of the IEEE Autotestcon*, 661-672.

84. Su, L.P., M. Nolan, and B. Norman (2000), Prognostics framework-Update II, *Proceedings of the IEEE Autotestcon*, 497-504.

85. Sumita, U. and J.G. Shanthikumar (1985), A class of correlated cumulative shock models, *Advances in Applied Probability*, **17**, 347-366.

86. Telek, M. and A. Heindl, (2003), Matching moments for acyclic discrete and continuous phase-type distribution of second order, *International Journal of Simulation*, **3**, 47-57.

87. Timm, N.H. (2002), *Applied Multivariate Analysis*, Springer, New York.

88. Tsoukalas, L.H. and R.E. Uhrig (1997), *Fuzzy and Neural Approaches in Engineering*, John Wiley and Sons, New York.

89. Valdez-Flores, C. and R.M. Feldman (1992), An improved policy iteration algorithm for semi-Markov maintenance problems, *IIE Transactions*, **24** (1), 55-63.

90. Virkler, D.A., B.M. Hillberry, and P.K. Goel (1979), The statistical nature of fatigue crack propagation, *ASME Journal of Engineering Materials and Technology*, **101** (2) 148-153.

91. Wackerly, D.D, W. Mendenhall III, and R.L. Scheaffer (1996), *Mathematical Statistics with Applications*, Duxbury Press, Belmont, California.

92. Waldmann, K.-H. (1983), Optimal replacement under additive damage in randomly varying environments, *Naval Research Logistics Quarterly*, **30**, 377-387.

93. Whitt, W. (1984), On approximations for queues, III: Mixtures of exponential distributions, *AT & T Bell Laboratories Technical Journal*, **63** (1), 163-175.

94. Wilson, B.W., N.H. Hansen, C.L. Shepard, T.J. Peters, and F.L. Greitzer (1999), Development of a modular in-situ oil analysis prognostic system. *International Society of Logistics (SOLE) Symposium*, Las Vegas, Nevada.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
14-09-2004	**Doctoral Dissertation**	Jun 2001 – Sep 2004

4. TITLE AND SUBTITLE

HYBRID STOCHASTIC MODELS FOR REMAINING LIFETIME PROGNOSIS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
2004-009

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Cox, Steven M., Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Street, Building 642
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DS/ENS/04-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Juan R. Vasquez, Major, USAF, Ph.D.
Air Force Office of Scientific Research
4015 Wilson Blvd, Room 173
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The United States Air Force is developing its next generation aircraft and is seeking to reduce the risk of catastrophic failures, maintenance activities, and the logistics footprint while improving its sortie generation rate through a process called autonomic logistics. Vital to the successful implementation of this process is remaining lifetime prognosis of critical aircraft components. Complicating this problem is the absence of failure time information; however, sensors located on the aircraft are providing degradation measures. This research has provided a method to address at least a portion of this problem by uniting analytical lifetime distribution models with environment and/or degradation measures to obtain the remaining lifetime distribution.

15. SUBJECT TERMS

Reliabililty, Life Expectancy (Service Life), Degradation,
Semi-Markov Processes, Markov Processes, Homogeneous, Nonhomogeneous, SMP, CTMC,
Prognosis, Remaining Lifetime Distribution, Lifetime Distribution, Phase-Type Distribution, Matrix-Exponential Distribution

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jeffrey P. Kharoufeh, AFIT/ENS
U	U	U	UU	178	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4603; e-mail: Jeffrey.Kharoufeh@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18